The Albelda Clayton-Matthews Paid Family and Medical leave Simulator Model Documentation: A report and simulation model presented to the Women's Bureau of the Department of Labor

Alan Clayton-Matthews  
*Northeastern University, a.clayton-matthews@neu.edu*

Randy Albelda  
*University of Massachusetts Boston, randy.albelda@umb.edu*

Follow this and additional works at: [http://scholarworks.umb.edu/econ_faculty_pubs](http://scholarworks.umb.edu/econ_faculty_pubs)

Part of the Economics Commons

Recommended Citation

[http://scholarworks.umb.edu/econ_faculty_pubs/41](http://scholarworks.umb.edu/econ_faculty_pubs/41)
The Albelda Clayton-Matthews Paid Family and Medical leave Simulator Model Documentation

A report and simulation model presented to the Women’s Bureau of the Department of Labor

by

Alan Clayton-Matthews and Randy Albelda

November 10, 2015

Alan Clayton-Matthews
Associate Professor
School of Public Policy and Urban Affairs
Northeastern University
Boston, MA 02115
a.clayton-matthews@neu.edu
617-373-2909

Randy Albelda
Professor
Economics Department and Public Policy Ph.D. Program
University of Massachusetts Boston
Boston, MA 02125
randy.albelda@umb.edu
617-287-6963

This workforce product was funded by a grant (WB-26510-14-60-A-25) awarded by the U.S. Department of Labor’s Employment and Training Administration. The product was created by the grantee and does not necessarily reflect the official position of the U.S. Department of Labor. The Department of Labor makes no guarantees, warranties, or assurances of any kind, express or implied, with respect to such information, including any information on linked sites and including, but not limited to, accuracy of the information or its completeness, timeliness, usefulness, adequacy, continued availability, or ownership. This product is copyrighted by the institution that created it. Internal use by an organization and/or personal use by an individual for non-commercial purposes is permissible. All other uses require the prior authorization of the copyright owner.
@ Copyright 2015 by Randy Albelda and Alan Clayton-Matthews. All rights reserved.
I. Introduction
The basic strategy behind our approach to estimating the cost of a paid leave program in Massachusetts was to, as much as possible, base estimates of program costs on actual known leave-taking behavior, and where this was not possible, to estimate a range of program costs reflecting a range of reasonable assumptions about unknown aspects of behavior in the presence of a paid leave program. We wanted to be able to estimate the sensitivity of program costs estimates to these assumptions. We also wanted to be able to analyze the distribution of program benefits by demographic characteristics. Furthermore, we wanted users to be able to estimate the costs of similarly structured paid leave benefit programs in other states, to be able to have some control over the assumptions about behavior that affect program cost estimates, and to be able to undertake their own distributional analyses.

We chose a simulation strategy as the best way to accomplish these goals. To obtain the best estimates possible about known leave-taking behavior, we use the Public Use Family and Medical Leave survey data collected by Abt Associates in 2012 for the Department Labor (referred to here as the DOL Survey) (McGarry, et al, 2013) to estimate behavioral models of leave-taking behavior conditional on the demographic characteristics of individuals, and use the Census Bureau’s American Community Survey Public Use Microdata Sample (hereinafter referred to as the ACS or ACS PUMS) to predict leave-taking behavior conditional on the demographic characteristics of individuals.

The DOL Survey is the best available source of information on leave-taking behavior. It is a representative national sample of leave takers, leave needers (those persons who said they needed but did not take a leave), and other workers who did not take a leave. The survey, which was conducted between February and June 2012, includes extensive information on the number and types of leaves taken, how long they were, whether and to what extent the employer provided pay while on leave, and whether or not some or additional pay while on leave would result in a decision to take a leave or to have taken a longer leave. The survey includes several demographic characteristics
related to leave-taking behavior, including sex, race and ethnicity, martial status, the
presence of children, education, family income, and whether or not the respondent was
paid on an hourly basis. The survey is used to estimate several aspects of leave taking
behavior, conditional on demographic characteristics and leave type. These include the
probability of needing a leave, of taking a leave, of getting paid for a leave, of extending
a leave if some or more pay were received, etc.

The ACS is a large national representative sample of persons. It is of sufficient
size to obtain reliable estimates of paid leave program costs and of the distribution of
program benefits at the state and sub-state level. The 5-year ACS PUMS can yield
reliably accurate estimates at geographic areas consisting of one or more PUMAs (a
PUMA is a geographic area that consists of a population of roughly 100,000 persons).
This survey also provides a rich array of demographic characteristics that closely match
those on the DOL Survey, which means that the behavioral models estimated on the DOL
Survey can be used to predict leave-taking behavior on the ACS.

The simulation model is a software application that “runs” each sample person
from the ACS through the estimated behavioral models and sets of assumptions about
leave-taking behavior. The flow of the person through the software mimics the sequence
of decisions and events that a person makes and experiences in the leave process. This is
an appealing aspect of simulation methodology since its structural approach helps
identify what assumptions are necessary in developing program cost estimates and at the
same time clarifies the impact of these assumptions on the bottom line estimates.

At several points during the simulation, such as when a person decides to take a
leave of a particular type or not, a decision is made based on a logit behavioral equation.
The logit equation estimates the probability of deciding “yes”. This probability, which is
a function of the person’s demographic characteristics, is compared to a random draw
from a standard uniform distribution (any point on the number line between zero and one
is equally likely to be chosen), and the random draw is compared to the probability given
by the behavioral equation. If the random draw is less than this probability (or less than
or equal, it really does not make any difference), the decision is “yes”, if not, “no”.

The
model flow then directs the person to the next point in the modeling sequence, depending on the result of this random draw. This is the essence of simulation.

After each person has been passed through the entire flow, the result is a history of leave-taking behavior for a one-year period. The model generates micro data output files consisting of records for each sample person and leave taken. These files can be analyzed with standard statistical software or database applications.

Aside from errors related to the DOL survey and estimates of the behavioral equation parameters, there are two sources of statistical error related to the simulator that are important to consider. One is sampling error due to the ACS. The ACS is a sample and is subject to sampling error that affects program cost estimates. The magnitude of this error is approximately inversely proportional to the square root of sampling size, and can be reduced by concatenating successive years of the ACS together. The second source of statistical error is due to the simulation methodology itself when the dependent variable is binary (or categorical). Even if the coefficients of a behavioral equation are “correct”, individual predictions are not at the individual level. For example, suppose a logit equation predicts that the probability of taking a leave is 30 percent for a person with a certain set of demographic characteristics. For any single person, the simulation results in either the person taking the leave – a simulation error of 70 percent-- or the person not taking the leave – a simulation error of 30 percent. The law of large numbers assures that the error approaches zero on average as the number of persons “run” through this equation approaches infinity. The magnitude of this simulation error is inversely proportional to the square root of the number of “runs” through the equation. The incidence of some types of leave is small enough that this source of error is not negligible. This type of error can be reduced by concatenating ACS data files, but there is also another way to reduce simulation error. That way is to “clone” the sample ACS person (i.e., to create several duplicates of the same person) and to run each duplicate person through the simulation. The software allows the user to specify this option. At the state level, the ACS sample is large enough so that cloning is not really necessary; but for estimates at sub-state geographies, cloning may be an excellent way to reduce simulation error.
The next major section of this document describes the basic modeling strategy used by the simulation model. The third major section of this document describes how to use the simulation model. It contains three subsections. The first gives instructions on how to install the software on a computer, and describes the organization of the folders and what they contain. The second describes how to run the application and how to form the commands that control the model’s parameter “handles” and direct the model’s input and output files. These parameters allow the user to change the paid leave program’s specifications, and also to change assumptions about participation and leave extension behavior. The third describes the output files created by the model and how to use statistics or database software applications to access these files. The fourth major section describes the flow of the model.

II. Modeling Strategy and Assumptions

The principal strategy behind the implementation of the model is to use econometric estimates of known leave-taking behavior when possible, and to incorporate reasonable assumptions and user-supplied options about unknown behavior. As new knowledge about behavior becomes available, the user may be able to incorporate that in model options – for example, new knowledge about take-up rates. In addition, knew knowledge may be incorporated as it becomes available in future versions of the model.

Modeling Known Behavior

The best source of information on which to model several aspects of known behavior – the incidence of taking or wanting to take a leave of a particular type, the probability of receiving pay while on leave and the amount of pay received, the length of leaves taken, and the probability of meeting the eligibility requirements of a proposed paid leave program – is the Family and Medical Leave in 2012 survey by Abt Associates for the Department of Labor (McGarry, Klerman, Daley, and Pozniak, 2013). The population surveyed consisted of adults 18 and older who had worked for pay in the last 12 months. They were asked about leaves taken or wanted during the prior 18 months for reasons of own health or disability (including maternity disability); to care for a new child; for health conditions of children, spouses, parents, other relatives, and non-relatives; and for issues arising from the deployment of a military member. Due to small
sample sizes for some categories, we limited our analysis and modeling to the following six leave types:

1. Own health;
2. Maternity disability;
3. Care for a new child;
4. Ill child;
5. Ill spouse; and
6. Ill relative.

The sample of persons surveyed can be classified into four groups depending on whether they took a leave or wanted to take a leave or not:

1. Those who took a leave and who did not not take a leave they wanted to take (leave takers only, N=1,133);
2. Those who wanted to take a leave but did not take any leaves (leave needers only, N=219);
3. Those who did not take a leave or want to take a leave (employed only, N=1,301); and
4. Those who both took a leave and also did not take a leave that they wanted to take (dual takers/needers, N=199).

The sample was weighted to the population so population rates and totals could be inferred from the sample. The survey asked about the longest and most recent leaves taken or wanted – and the reason for that leave – in the last 18 months, whether those leaves were taken or wanted in the last 12 months, and how many leaves in all were taken in the last 18 months and in the last 12 months. Leaves were counted by “reason”, so intermittent leaves for a single reason were counted as a single leave. Leave takers were asked about the reasons and lengths of leave for up to two leaves: the longest and the most recent (often they are the same). Leave needers were asked about the most recent leave needed and the reasons for up to two more leaves needed. For both taker and needer
leaves, respondents were asked if they saw a doctor or had a hospital stay. For the most recent leave taken or needed, additional information was asked. For leave takers, this included questions about pay received while on leave; and if full pay was not received, whether they would have taken a longer leave if they had received additional pay. For leave needers, this included a question about why they didn’t take the leave. Many respondents volunteered that they couldn’t afford to take an unpaid leave. These questions about additional pay and affordability were helpful in modeling the response of leave lengths and participation in the presence of a paid leave program. Leave takers were also asked about whether some of the pay received while on leave was part of a TDI program or a state family leave program.

Respondents were also asked about their work. Particularly useful for modeling behavior and estimating program eligibility were questions about weekly hours, whether they worked full year and were continuously employed by a single employer, how many employees worked at their organization within 75 miles, and whether they were paid on an hourly basis or not. Demographic information on respondents included age, sex, race/ethnicity, marital status, educational attainment, family income, and how many children were in their care.

Earlier theoretical work and statistical analysis of a prior Department of Labor family leave survey (Westat, 2001; Albelda, and Clayton-Matthews, 2010) established that the information in this survey would be useful in estimating statistical models of the probability of taking or needing a leave of a particular type, the probability of receiving partial or full pay, and the probability of meeting the eligibility requirements of the FMLA law or a proposed paid leave program. The estimation strategy involved a specification search that began with a full set of demographic and economic variables and “tested down” to a specification that included independent variables that were at or near statistical significance at the 5% level and that “made sense” in terms of yielding estimated coefficients of the expected sign and reasonable magnitude. The estimated relationships appear in the model source code (in the Parameters class) and may be reported in a separate publication.
These statistical models are implemented in the model by applying the estimated coefficients to variables on the ACS for each sample individual worker. Most of these models estimate a probability: the probability of taking or needing a leave for a particular reason, the probability of receiving pay while on leave, the conditional probability that that pay was full pay, etc. Using the coefficients of the logit regression model and applying them to the sample individual’s independent variables yields a probability of taking or needing a leave, of receiving pay while on leave, of receiving full pay conditional on receiving any pay, etc. These probabilities are compared to a random draw from a standard uniform distribution – using the model’s pseudo random number generator – to determine whether an outcome happens or not. Other models – for example, for fraction category of pay received, or number of leaves taken – are estimated by an ordered logit model, and the random draw determines the category by the estimated cumulative probability distribution of outcomes. Several models – usually when sample sizes are too small to estimate probabilities conditional on observable characteristics – are simply the weighted distributions from the survey. These models are identical to statistical models that contain only a constant, and are handled by the simulator in the same manner as other models that predict probabilities of binary outcomes or ordered outcomes.

This strategy “works” because the both the DOL survey and the ACS are representative samples (after weighting) of the population and both contain closely similar measures of independent variables. The match of variables is not complete, however, so a few variables not available on the ACS have to be imputed. For behavioral models these involve two variables: whether or not the worker is paid on an hourly basis, and whether or not the worker is covered and eligible under the FMLA law. This eligibility criterion is significant in several behavioral relationships, and involves weeks worked, worked full time continuously for a single employer in the past 12 months, and worked for a firm that had at least 50 employees within 75 miles. Other eligibility requirements of proposed paid leave programs might require knowing weeks worked; and benefit rules of proposed programs usually pay benefits proportional to weekly earnings.

The ACS does not ask whether pay is received on an hourly basis; does not ask about employer size; does not ask about the number of employers that the person worked
for in the last 12 months; does not ask about weekly pay; and records weeks worked in aggregated categories. These variables are imputed on the ACS using models and distributions estimated from the Current Population Survey (CPS) – which does include these variables – conditional on demographic and economic variables common to both the CPS and ACS surveys. These models and their estimated parameters are included in the Wage class in the software. Weekly wages on the ACS are estimated as annual earnings divided by the imputed number of weeks worked.

*Simulating Unknown Behavior*

Some information about leave-taking behavior needed for our simulation procedure cannot be estimated from the DOL 2012 survey, although some information collected there is useful in making some reasonable assumptions. The three main pieces of unknown information – whether a worker will use a paid program or employer benefits; program take-up rates; and whether a worker will extend a leave in the presence of a program – are discussed below.

1. *How employer benefits affect participation in paid program.*

The decision to participate in the paid leave program, given that a person is eligible, will in large part be based on the level of program benefits the worker would receive compared to the next best alternative. These alternatives consist of employer pay (if the person receives it) or nothing (if the leave is unpaid in the absence of the program). In order to compensate for the time and effort of applying to the program, program benefits would have to exceed the next best alternative by some amount. This amount may differ systematically by income and by other factors. It may also vary randomly across different individuals, and even for the same individual, at different times.

In the model, this participation decision is implemented by an arbitrary logit equation with two independent variables: the difference between weekly paid program benefits and weekly pay received while on leave, and family income. The participation probabilities it yields are given in Table 4 for several combinations of benefit/pay differentials and family income.
2. Take-up Rates

The simulation model estimates the number of all eligible workers that would use a paid leave program in light of current employer benefits. This estimate assumes that everyone taking a leave knows about the program and that the program is virtually costless to use. That is, the output from the simulator assumes a 100 percent take-up rate. However, this is completely unrealistic, which is why one of the policy parameters that can be adjusted by the user is take-up rates. The degree to which eligible leavers might use a paid leave program depends on a variety of factors beyond the scope of what can be uniformly modelled or assumed. Three important ones are: general knowledge of the program by workers; administrative complexity in obtaining program benefits; and workplace culture that either encourages or inhibits use. Recent experiences with care and bonding leaves in California, New Jersey and Rhode Island suggest that take-up rates, at least for several years, will be low. A recent estimate indicated that 25-40 percent of new mothers used the 6 week care leave in California, even after 10 years of implementation. Appelbaum and Milkman found that fewer than 50 percent of California workers knew about paid family leave. The degree to which state administrators and paid FML advocates work to make the program known will positively affect take-up rates. Use of any program will require time on the part of leave takers (and employers) to fulfill the administrative requirements of the leave. An easy-to-use program can reduce that time (which is a cost). Still, workers that take relatively short leaves may not bother at all. There may be other real or perceived costs to taking a program leave. If workers fear their position at their job might be threatened if they take a leave, then take-up rates will be low. For example low-wage workers may fear being replaced altogether while high-wage employees may fear an employer might not provide them with better opportunities.

The simulation model lets the user select a take-up rate based on some reasonable assumptions about the percent of eligible workers that might use a particular program in a particular state. The user can apply different take-up rates for different kinds of leaves. For example, there may be reason to believe that maternity disability leaves might have higher take-up rates than other leaves. Almost all mothers that give birth do leave work for a continuous period of time that is usually known in advance. Employers and employees typically expect new mothers to be away from work for more than a few
weeks. Further, obstetricians and others in pregnant mother’s networks are likely to inform them of a paid leave program so usage might be higher than other types of leaves.

3. Extending a leave in the presence of a program

In the presence of a paid leave program, leaves would not be shorter than in the absence of the program, but they may be longer. Lacking empirical evidence about the effect of program benefits on extending leave lengths, we estimate the probability of extending a leave. Because this decision is complex and affected by length of leave before the decision to extend, availability of employer pay, and whether the leave is job-protected, we use different extension rules in the simulation. For workers with short leaves (leaves that end before the waiting period of the program is over), we estimate the probability of taking a longer leave using logit regression estimations relying on the response to the DOL survey question, “Would you take a longer leave if you received some/additional pay?” If the model simulates an extension, we arbitrarily extend the leave for 1 week. We assign a different decision to those employees who reach the end of their original leave length (the length they would take if there was not program) and are receiving either program or employer benefits (but not both). We assume the probability of extending these leaves using program benefits are 25 percent and for those who do extend, that the extension is equal to 25 percent of their original length, not to exceed the maximum length of the program. The last decision applies only to those who have exhausted the paid program and still have some employer benefits available to them (based on the simulation). In this case the simulator assigns them a 50 percent probability of taking an extended leave for as long as they still have employer benefits. In all cases, if the original length of leave is less than the FMLA job-protection length of 12 weeks, an option in the model allows the user to restrict the leave to a maximum of 12 weeks. In the case of own health leaves, there was a significantly longer distribution of leaves for workers who received some part of their pay from state programs; and so the model incorporates longer own health under a paid leave program by using the distribution of leaves experienced by these covered workers.
III. How to Use the Model

**How to Install the Model**

The model comes in a two-level folder hierarchy. The main folder, named “FML2” (called the FML folder) contains the simulation executable programs and subfolders that contain the documentation, software, parameters, input and output files. These are named “Documentation”, “Software”, “Parameters”, “Input”, and “Output” respectively.

Copy the FML folder and its contents to your computer’s internal hard drive. It can be placed anywhere in a user’s directory structure that is convenient. It is important to maintain this relative directory structure, since the relative paths for the input, output, and parameter files are “hard coded” into the software.

**System Requirements**

Windows 7 or later operating system with Service Pack 1, Microsoft Visual Studio Express 2015 for Windows Desktop installed (a free download from Microsoft is at: [https://www.visualstudio.com/products/mt238358](https://www.visualstudio.com/products/mt238358)), 8GB RAM, 30GB of hard drive space. The Visual Studio Express 2015 is needed because it installs certain required operating system files. The hard drive space needed depends on how many ACS input files you need. Thirty GB is enough to hold the 5-year U.S. ACS PUMS 2009-2013 data set as well as a few state ACS data sets.

**The Directory Structure**

The executable programs are in the FML folder. This version of the model contains two executables: “FML2_State_2009-2013.exe”; and “FML2_US_2009-2013.exe”. The former runs on a 5-year 2009-2013 ACS PUMS dataset, and the latter runs on the 5-year U.S. 2009-2013 ACS PUMS dataset. The subfolders and their contents are described in outline form below.

**Documentation**: This folder contains this documentation, the command reference appendix as a separate document, and the PUMS data dictionary from the U.S. Census Bureau.
**Input:** This folder contains the ACS files that serve as input files for the simulator, as well as the command and other user-created input files described in the command reference appendix. This folder may already contain the 5-year U.S. ACS PUMS 2009-2013 files (8 files in all in .csv format) and/or a state 5-year ACS dataset. Download any 5-year ACS PUMS 2009-2013 files you would like to use directly from the Census Bureau to this folder. The model uses the “.csv” formatted files from the Census Bureau, which can be obtained at [http://www.census.gov/programs-surveys/acs/data/pums.html](http://www.census.gov/programs-surveys/acs/data/pums.html). This link is also available from the American Community Survey’s home page [http://www.census.gov/programs-surveys/acs/](http://www.census.gov/programs-surveys/acs/). From this home page, choose the “Data” link, and then the “PUMS Data” link. There may be an example command file (cmd.txt) included in this input folder, as well as other user-supplied sample files that are referenced in the command file.

**Output:** This folder will contain the output files produced by the simulation. If you plan to run several simulations and you want to keep a record of what you did, it is recommended that you manage your output files, perhaps by copying them after each run into a subfolder of this folder, or to some other place.

**Parameters:** This folder contains several text files of parameters that are read by the simulator when it starts up. Do not change these files! It is recommended that you do not touch this folder. If you would like to read or open the files in this folder, it is recommended that you change the files’ properties to “ready only” so you do not accidentally change them.

**Software:** This folder contains the source code for the simulator. These files can be opened and inspected in any text editor. They are not used by the simulator, so you can alter them – for example, by putting in comments – as you wish. This software contains the code for the simulator’s procedures, so that an experienced C++ programmer can “look inside” the simulator’s engine. However, it does not contain the code for the objects and containers used by the model, for example, the code that implements containers like matrices and arrays, objects that handle dates, code that parses commands, etc.
How to Run the Model

To start the simulator, double click the “FML2_State_2009-2013.exe” or the “FML2_US_2009-2013.exe” file icon in the FML folder. Your system may not display the “.exe” extension. The simulator will open a window and prompt you for the command file name. Include its extension (“.txt”) as well.

The command file tells the simulator what input and output files it will use, what level of detail should be output, what ACS variables to include in the output, whether cloning should be used, what the program eligibility requirements and program benefit rules are, and what options the simulator should use that affect program participation and leave lengths. All input files should reside in the Input folder before the simulator is run. All output files will be written to the Output folder, over-writing any files of the same name. The commands, what they do, their syntax, and examples are given in the Command Syntax appendix. A sample command file is given in Appendix A.

How to Analyze the Model Results

The output files that contain results of the simulation are all in comma-separated format, with variable names in the first row, and so can easily be read by statistical and database software packages. Two files are also created to aid in processing the output. One is the documentation file, which contains information on all the variables in the output files. For Stata users, another file, called “labels.do”, is also created in the Output folder. It contains Stata label commands for each variable.

Identifier variables are output to each file to enable merging of information from one file to another. The “personid” can be used for merging the “main” and “leaves” files. The “personid” and “leaveid” variables are available on every file except “main” (since “main” is at a more aggregate level than a leave). The “nstate” variable can be used, along with “personid” and “leaveid”, to link the three state files together (the “states”, “weekly benefits”, and “weekly employer payments” files). Finally, “weekid” in the “weekly employer pay and program benefits” file, “benwkno” in the “weekly benefits” file, and “paywkno” in the “weekly employer payments” file all refer to the same week within each leave, allowing these files to be merged if desired. Appendix B
provides a complete list of variables names, label, level, and data source for each output file.

Each file also contains a variable called “weight” that should be used to weight records up to population totals. The variables from the ACS maintain the same spelling as in the official ACS data definition files, so those files, included in the Input folder, can be used for variable codes. Several date variables come in two forms: 1) a triplet of variables recording the month, day, and year; and 2) the number of days since January 1, 1960. This latter form is used by Stata, which makes handling of dates convenient in Stata.

The key variable for calculating aggregate paid leave program benefit costs is the variable “benamt” on the “leaves” file. The weighted sum of this variable gives total program benefit costs.

 Technical Support
For technical support, contact Alan Clayton-Matthews at a.clayton-matthews@neu.edu, 617-373-2909.

IV. What the Model Does: The Flow of the Model

This section first provides programming vocabulary and then describes what the simulation does by following the flow through the model’s software. For the most part, this flow corresponds to the timing of decisions and modeling of behavior individuals make and exhibit in the process of taking a leave for personal or family-related medical reasons. Again, the way in which this simulator models the leave process, including the simulated behavior and personal decisions, are highly influenced by, and constrained by information and the structure of the DOL survey.

 Some Programming Vocabulary: Classes, Objects, Instances, and Procedures

The simulator is written in the programming language C++ (C Plus Plus), an object orientated language that facilitates the construction of complex software applications by allowing the programmer to break the problem down into a small number
of “objects” called “classes.” Each class is more or less self-contained and corresponds
to a logical piece of the problem that usually represents a physical object (like a person or
accounting ledger book), or a concept (like a set of benefit program rules or behavioral
equations). A class is a template, consisting of source code, that describes the
characteristics of an object and how it behaves, and an object is a particular “instance” of
the class. The software code that describes the rules of how objects work is arranged into
units of code called “procedures”. These are sometimes also called “methods”; and a
Fortran programmer might call them “subprograms”. Each object of the same class has
the same procedures, yet each object can – and almost always has – a different set of
data. For example, each person who is run through the simulator is an object with
different characteristics than other persons. Each person might have none, one, or more
leave objects that describe the characteristics of each leave. Each leave might involve a
different sequence of events, where each event is a “state” object.

A note on syntax:

We refer often to the source code document and procedure or class within the
source code document. The syntax is as follows: the source code document file name
appears in italics and the procedure appears as regular type, with the two separated by a
comma. When the procedure belongs to a class, the class name precedes it, and is
separated from the procedure name by a double colon. The whole reference is enclosed
in parentheses. For example, (benefitcalc.cpp, BenefitCalc::MakeNamesBenefits) refers
to the procedure “MakeNamesBenefits” in the class “BenefitCalc” in the source code file
“benefitcalc.cpp”. Note that all source code files end in either the extension “cpp” or “h”.
Input and output file names will also be italicized.

Initialization of Program Parameters and Behavioral Relationships

After the application prompts for file names and user-controlled program
parameters (which are handled in (fml2.cpp, main)) the application creates an instance of
the class named “Parameters”. There are three distinct sources the simulator uses to
generate the information contained in “Parameters” class. The first source is user-
supplied program parameters (provided by the user and include the specifications of the paid leave programs).

The second source of information in the “Parameter” object comes automatically from files generated from tabulations of the DOL Survey. These file-read parameters are the distributions of leave lengths. The names of these files are listed in Table 1 and each can be found in the Parameters folder of the FML software application. Each type of leave has a different distribution of lengths, and except for own health, are different for men and women. Each file gives the cumulative distribution of leave length in days. The first row of each table gives the dimensions of the distribution matrix. The following rows contain the cumulative distribution. The first column gives the number of days, and the second column gives the proportion of leaves whose length is equal to or less than the corresponding number of days. These distributions are tabulated from the DOL Survey. Table 2 contains the information in the file “FML/Parameters/length NEW CHILD women 2.txt.”

Analysis of leaves lengths using the DOL Survey indicates that leave lengths of illness types are related to the severity of illness. However, aside from the gender of the leave taker (for all but own-health) and severity of illness, there are no other significant predictors of leave length. Importantly, whether or not the leave taker receives pay from his/her employer does not seem to be associated with the length of the leave. Since the ACS does not have information on individuals’ illnesses, the application simulates leave length by randomly drawing from the distribution that corresponds to the type of leave and gender of the leave taker.

The third source of information is initialized by the simulator when it creates the parameter object. This information concerns the amount of pay received while on leave. These estimates are from the earlier 2000 DOL Survey (Westat, 2001), since the newer survey did not ask these questions. Those leavers who indicated that they received partial pay from their employer while on leave were asked if they received at least some pay for each pay period that they were on leave (HA10D), and if not, was the pay for their full salary or only for a part of their salary (HA10E). Leavers were also asked what proportion of usual pay they received in total over the entire length of the leave (HA10F).
The responses to these questions were tabulated separately for each leave type, and expressed as conditional probabilities. They appear in (params.cpp, Parameters::Parameters). As an aid to reading these parameter values, the numbers for own health leaves are given in Table 3.

The Parameters class also contains the behavioral equations that estimate the probability of various events occurring, such as taking a leave, receiving pay, participating in the program, extending one’s leave in the presence of the program, etc. These are described below as they are encountered in the flow.

The Main Program Loop

The application reads the ACS input file household by household, and within each household, passes each person through the simulator (fml2.cpp, main).

First, it is determined whether or not the person is an adult civilian who worked last year, and was not self-employed (fml2.cpp, main; fml2.cpp, filterRequirements). Only these persons are in the universe of possible leave takers and passed through the rest of the simulator.

Some necessary information is not directly available on the ACS, and therefore is estimated or simulated. These include weeks worked (imputed from the categorical weeks worked variable), weekly wage (annual earnings divided by weeks worked), paid hourly or not, employer size, and worked for a single employer last year or not. The models for these imputations are in the Wage class (wage.cpp, Wage::Wage), except for employer size, which is imputed in (params.cpp, Parameters::EmployerSize).

Based on these imputations, the simulator next determines the work and employer-size eligibility requirements for FMLA and for the paid leave program, using information on the person’s work history. To approximate the work requirement under FMLA, the person had to have worked at least 1250 hours last year, and only have had one major employer last year (parms.cpp, Parameters::EligibleWorkerFMLA). In addition, for eligibility coverage under FMLA, the size of the establishment must be at least 50 employees (params.cpp, Parameters::EligibleEmployerWorkerFMLA). This concept of FMLA eligibility under the work and employer size requirements is used as an
independent variable in several of the behavioral equations in the model, because it influences the person’s ability and willingness to take a leave, and also is correlated with other personal and job characteristics that are not measured by other independent variables. Worker eligibility and employer coverage under the proposed program is calculated according to user-supplied eligibility requirements (params.cpp, Parameters::Eligible).

The person then enters the main software program loop illustrated in Figure 1. Each person is run through two branches illustrated in the figure (fml2.cpp, set_leaves). The person might be a leave taker, a leave needer, or both in a given year. On the left branch, the probability of a person’s most recent leave being each of the six possible leave types is estimated conditional on the person’s characteristics (params.cpp, Parameters::PrTake). These probabilities are compared to a draw from a standard uniform probability distribution. (Think of a “Wheel of Fortune”, where the size of each slice on the wheel is proportional to the probability of a particular leave type, with the remaining large slice representing no leave.) Note: except where noted in Figure 1, each arrow represents a positive outcome. A negative outcome results in the person “dropping out” from taking a leave.

If one of the leave types is chosen, the possibility of more than one leave is simulated (params.cpp, Parameters::PrMultipleLeaves); and if so, the number of leaves greater than one is simulated as a random draw from the probability distribution of 2 through 6 possible leaves (params.cpp, Parameters::PrDistMultipleLeaves). The types of these additional leaves, if any, is simulated from an estimate of the conditional probability distribution of a second leave (conditional on a first leave) (fml2.cpp, set_leaves). This conditional probability distribution was estimated from those sample persons in the DOL survey who reported on the type of leave for both their longest and most recent leave, when these were different leaves. The survey implied that the probability of taking a second ill child leave or a second ill parent leave was higher than the unconditional probability of each, and the probability of taking a second maternity disability or new child leave in a given year was effectively zero.
The leave length is simulated as a random draw from the estimated distribution of each type of leave length given by the DOL survey. Except for own health leaves, these differed by sex, with women tending to take longer leaves than men. For own health leaves, leave lengths were longer for those who stated that they received some pay from a TDI or state paid leave program, so two leave length distributions were used: in the absence of a program, the distribution of leave lengths for persons who did not report receiving these payments was used; in the presence of program, the distribution of leave lengths for persons who did report receiving these payments was used.

The text files that contain these estimated distributions are noted in Table 1. Leave lengths are counted in days, ignoring weekends, so a leave of two weeks, for example, is ten days. At this point in the program flow, the leave lengths represent those in the absence of a paid leave program, except for those persons who would not have taken a leave in the absence of such a program. Later in the flow, in the presence of the paid leave program, the person may choose to extend their leave.

Up to this point, the simulation on the right branch, for leave needers, is similar (fml2.cpp, set_leaves), except that simulated leave lengths represent leave lengths if they were to take a leave. The models used on this branch include (params.cpp, Parameters::PrNeed), (params.cpp, Parameters::PrMultipleNeeds), and (params.cpp, Parameters::PrDistMultipleNeeds).

For leave takers, their weekly payments while on leave in the absence of a program is simulated in stages. First, whether or not they receive any pay while on leave (fml2.cpp, processPerson) and (params.cpp, Parameters::PrPaidLeave). Next, conditional on receiving pay, was it full pay (fml2.cpp, benefits), (fml2.cpp, payGroup), (params.cpp, Parameters::PrFullPay); and if not, what fraction of pay was received (params.cpp, Parameters::OProbPayGroup).

For those who were partially paid, the 2000 DOL Survey asked if the respondent received some pay for each pay period that they were on leave; and if not, in the pay periods for which they did receive pay, was it for their full salary? As described in the section on parameter initialization and illustrated in Table 3, the survey was used to estimate these conditional probability distributions for each leave type and payment.
group (less than half pay, about half pay, more than half pay). If a person’s leave was partially paid, their payment schedule was randomly selected from the corresponding conditional probability distribution for their leave type (fml2.cpp, benefits; params.cpp, Parameters::PrSomePayCells).

At this point, the application has determined if a person received some pay each week; and if not, if that person received full pay for some weeks; and if, over the course of their leave, a person received less than half of full pay, about half of full pay, or more than half of full pay. The weekly pay schedule is then filled out using arbitrary rules subject to these payment schedule and amounts constraints. For example, those persons who received some pay for each week of their leave, but who received less than one quarter pay in total, were assigned 12.5 percent of their weekly pay in each week of their leave (fml2.cpp, benefits), while those persons who received some pay each week, and more than three-quarters but less than full pay, were assigned 87.5 percent of their weekly pay in each week of their leave (fml2.cpp, benefits).

For leave needers, the model simulates whether they would take a leave if there were a paid leave program based on their reason for not taking a leave being that is was not affordable (params.cpp, Parameters::PrTakeGivenProg). If not, they are classified as an ultimate leave needer. If they do, they then follow the same remaining path as leave takers.

At this point, the leave taker’s (or potential leave taker, if originally a needer), eligibility is determined (fml2.cpp, set_leaves). The work and employer eligibility conditions have already been determined by this point, so here it is determined whether or not they saw a doctor or went to a hospital (or whether the person they took a leave to care for saw a doctor or went to the hospital). These are computed by comparing the probability of a logit behavioral equation for each condition, i.e., seeing a doctor and going to the hospital (params.cpp, Parameters::PrDoctor or params.cpp, Parameters::PrDoctorNeed; params.cpp, Parameters::PrHospital or params.cpp, Parameters::PrHospitalNeed) to a corresponding random number. The doctor and hospital requirements vary somewhat depending on the leave type. Essentially, to be eligible for an FMLA-defined leave (except for new child) requires either seeing a doctor
or going to the hospital (*params.cpp*, Parameters::EligibleDoctorHospital), and it is presumed that if the person or the person they were caring for went to the hospital, they also saw a doctor.

After it has been determined what leaves, if any, the person takes, and their lengths, the leaves are then distributed across a calendar where their leave either finishes in a 12-month period beginning April 16, 2011 and ending April 15, 2012, or they are still on leave on April 15, 2012. The dates for the beginning and ending of the 12-month period are not critical. This period was chosen simply because the survey was conducted between February and June of 2012, so April 15 was approximately in the middle of this period. The possibility of having an unfinished leave at the end of the calendar is simulated by (*params.cpp*, Parameters::PrUnfinished). The random assignment to the calendar is given by (*Calendar.cpp*, Calendar::AssignLeaves). This assumes that leave-taking is not seasonal. Although the model simulates leave ending dates that are uniformly distributed throughout the year, it does not guarantee that for any simulated person, the dates “make sense” in that it is possible that two simulated leaves overlap in time. However, what it does achieve is a reasonable estimate for the extent to which some leaves which take place during a given year “spill” outside the yearly time period, either because they began before the year began or ended after the year ended.

*Employer pay, program benefits, and leave length in the presence of a paid leave program*

The next step in the model is to simulate employer pay, program benefits, and possible extensions of leave length in the presence of a paid leave program. The application simulates the sequence of events and choices that a leaver would reasonably experience, given their weekly leave history and weekly schedule of employer payments simulated up to this point, in the absence of a paid leave program. Three important, and reasonable, assumptions are embodied in this part of the simulation:
1. Not all eligible recipients participate in the program, due to lack of information, the hassle of applying, or other reasons. The proportion of eligible paid program leave takers who do participate is called the “take-up rate”, and is one of the program parameters set by the user. The user can specify a different take-up rate for each leave type.

2. The decision to participate in the paid leave program, given that the person is eligible, is also based on the level of program benefits he/she would receive compared to the next best alternative, which is employer pay if the person received it, or nothing, if the leave was unpaid in the absence of the program. In order to compensate for the time and effort of applying to the program, program benefits would have to exceed the next best alternative by some amount, and furthermore, this amount may differ systematically by income and by other factors. It may also vary randomly across different individuals, and even for the same individual, at different times. This participation decision is implemented by an arbitrary logit equation that has the difference between weekly paid program benefits and weekly pay received while on leave, and family income as independent variables. The participation probabilities it yields are given in Table 4 for several combinations of benefit/pay differentials and family income. The user can turn off this participation module if they wish, and rely solely on the take-up rate parameter instead (see section How to Run Model).

3. The part of the simulation flow that establishes employer pay, program benefits, and possible extensions of leave length in the presence of a paid leave program is illustrated in Figures 3 through 5. The software models the process as a sequence of states (i.e. section of source code with at least one path of entry or exit), represented as circles in the diagrams. Some of these states are “decision” states, in which the leaver must make a decision to participate in the program, or to extend their leave beyond its “original” length, the length simulated prior to this point. The transition from state to state, represented by arrows, is the result of events or decisions, such as the end of receipt of employer pay, the original length of leave being reached, the
decision to participate or extend a leave, etc. Diamonds represent predetermined conditions or conditions over which the person has no control, such as whether the person is eligible or receives employer pay. The flow from state to state traces a “path”. Table 5 lists each of the states as used in the source code referred to in Figures 3-5.

In order to illustrate how this part of the simulator works, let’s follow one of the many possible paths that a leaver might experience. Suppose that the person originally took a partially paid leave of 3 weeks from her employer, got paid 30 percent of her full time pay for the first two weeks of leave, and is eligible for the paid leave program. This person might take the following path.

The person beings in state number 0, “Begin employer leave, eligible for paid leave program” (*fml1.cpp*, benefits; *benefitcalc.cpp*, BenefitCalc::BenefitCalc; *benefitcalc.cpp*, BeginPaidLeaveElig::BeginPaidLeaveElig).

After the waiting period of a week, the person decides whether or not to participate in the program, and is in state number 8, “Participate in paid leave program?” (*benefitcalc.cpp*, Participate_1::Participate_1). The weekly benefit from the paid leave program substantially exceeds the person’s partial pay from her employer, so that the probability of participation in the program is high (*params.cpp*, Parameters::PrParticipate1). In fact, suppose that the probability is greater than the random number drawn, so that the person participates.

The person is then in state number 2, “Begin paid leave program” (*benefitcalc.cpp*, BeginProgram::BeginProgram).

After being on the program for two weeks, the original leave length is reached, and the person then decides whether or not to extend her leave. The person is in state number 12, “Extend?” (*benefitcalc.cpp*, Xtend_1::Xtend_1). The simulation sets the probability of extending the leave to 25 percent. (*params.cpp*, Parameters::PrExtend1). Suppose the random number drawn is less than .25, so the person extends her leave. The
leave is then extended by 3 days, or 25 percent of the original leave length of 15 days (params.cpp, Parameters::DaysExtend).

The person is then in state number 7, “Begin/continue paid leave program” (benefitcalc.cpp, ContinueProgram::ContinueProgram). The person remains in this state for three days, and then ends her leave, entering state number 20, “End leave” (benefitcalc.cpp, EndLeave::EndLeave).

The “End leave” state in the simulator performs some accounting and cleanup tasks (benefitcalc.cpp, BenefitCalc::cleanup). Among other tasks, a number is assign to a variable called “path” that uniquely identifies this path that the person traversed. The “path” variable is calculated as the sum of a number of terms, where each term is the number two raised to the power given by the state number. The “End leave” state is omitted from this calculation since every leave ends in the “End leave” state. The value of “path” calculated for this particular path is 4485.

The processing for each person ends with outputting nearly all the information from the simulation to several files. Information from each leave is output to the “leaves” file, which contains the path traversed for each leave, along with summary information on program benefits received, employer pay received, the type of leave, its length in days, and other information. A record of the traverse through each state is also output to the “state” file, and the weekly record of program benefits and employer payments are written to three “weekly” files. The “main” file outputs information at the person level, including a summary of the leaves the person took, if any, and demographic information from the CPS so that the distribution of program benefits can be analyzed.

V. Conclusion
The application provides a sophisticated estimation of family and medical leaves. The program is relatively user friendly and includes considerable flexibility for the user in terms of possible paid leave program specifications and behavioral responses to a paid program. The application is a powerful estimation tool that can be used at the national, state, or even the metro area or city level. It not only estimates the amount that a paid leave program might cost, it also estimates the amount of employer benefits paid (in the absence and presence of a paid leave program) as well as employee uncompensated
wages due to any portion of earnings that are unpaid during a leave. Further, because the simulator outputs add the simulated variables onto the ACS, it is possible to look at the beneficiaries of paid leave programs by gender, race and ethnicity, marital status, and income level.
References


