Worker Paid Leave Usage Simulation (Worker PLUS) Model

Issue Brief: A Benchmarking Study of the Worker PLUS Model Results

January 2021

OVERVIEW

To facilitate understanding of the potential impacts of different policy alternatives on workers' leave-taking behaviors and program costs, the U.S. Department of Labor's Chief Evaluation Office contracted with IMPAQ International, and its partner Institute for Women's Policy Research (IWPR), to develop the Worker Paid Leave Usage Simulation (Worker PLUS) model, an open-sourced microsimulation tool based on public microdata and predictive modeling. The model and other relevant materials are publicly available at [*hyperlink*].

In this issue brief, we provide a benchmarking study of the model simulation results. The results from the Worker PLUS model are compared to those from an existing paid leave simulation model developed by Albelda and Clayton-Matthews (2017, the ACM model) and actual program administrative data. Simulation results compared include program benefit outlays and program participation for three state paid leave programs in California, New Jersey, and Rhode Island. The goal of this study is to provide users of both models a summary of model validity based on actual program data. The study shows that, although both models can predict program outlays fairly accurately, the Worker PLUS model improves upon the ACM model by better predicting program participant counts and by providing alternative estimates through multiple simulation methods.

This issue brief provides a benchmarking study of the Worker Paid Leave Usage Simulation (Worker PLUS) model developed by IMPAQ International (IMPAQ) and the Institute for Women's Policy Research (IWPR) under a contract with the U.S. Department of Labor (DOL). Results from the Worker PLUS model are compared to the existing Paid Family and Medical Leave Simulator model developed by Albelda and Clayton-Matthews (2017, the ACM model)[i] as well as actual program data from existing state paid leave programs. Through this study, users of the ACM model and the Worker PLUS model will be able to identify and reconcile similarities and differences between the two models and evaluate the predictability of model simulations.

The ACM model was initially developed in 2007 and has been revised and updated several times in intervening years, with the most recent update occurring in September 2017. Similar to the Worker PLUS model, the ACM model trains predictive models of individual-level leave-taking behavior using data from the U.S. Department of Labor Family and Medical Leave Act (FMLA) Employee Survey, and applies the trained models on the American Community Survey (ACS) Public Use Microdata Sample (PUMS) to simulate individual-level leave-taking behavior for state samples.[ii] The ACM model, however, has the limitations of (i) adoption of many heuristics, such as simulating leave lengths that are unconditional on leave reasons and applying take-up rates that are not validated against program administrative statistics; (ii) not offering a graphical user interface (GUI); (iii) lack of alternative simulation method options besides the traditional logistic regression model; and (iv) lack of transparency in simulation code due to the proprietary nature of the tool and the underlying C/C++ programming language, which can be difficult for modern data scientists and policy analysts to interpret.

The Worker PLUS model offers several enhancements to the ACM model: (i) a more rigorous simulation engine, with simulated leave lengths conditional on leave reasons and take-up rates calibrated against program administrative statistics; (ii) improved structures of model output and an easy-to-use GUI; (iii) shorter runtime; (iv) simultaneous comparisons across multiple simulations under a single model execution; (v) options of both traditional and machine learning simulation methods; and (vi) open-source coding in both Python and R, two of the most popular modern languages among data scientists, to allow for greater transparency and flexibility to users and researchers.[iii]

We test both models' predictive performance by comparing the model results against administrative program statistics reported from existing programs in California, New Jersey, and Rhode Island between 2012 and 2016, the latest period compatible with the ACM model developed in 2017. Outcomes compared include program benefit outlays, total eligible worker counts, program participant counts, and participant leave lengths. Program administrative and budget financing costs are not considered in these analyses, as the feature of calculating these costs is not available from the ACM model.[iv]

METHODOLOGY

We perform two different types of comparisons in this issue brief:[v]

1. ***Comparing simulated and published program benefit outlays.*** Predicting benefit outlays is one of the primary uses of the models.[vi] The simulated benefit outlays are computed as the population-weighted sum of simulated benefits received by each worker in the ACS sample. For each worker in the sample, the benefit is simulated based on program eligibility, program

participation status, wage, and leave length. Therefore, by first benchmarking the outlay estimates, we can assess the validity of the models at the highest level.

2. ***Comparing simulated and published population statistics.*** The benchmarking is then performed for the lower-level intermediate outcomes that are used to compute benefit outlays, including:
   - Total number of workers eligible for the program
   - Total number of program participants
   - Average length of program-paid leaves

   The benchmarking is not performed for wages because wage data are not available from state programs. The benchmarking analyses for program participant count and leave length are categorized by the six major leave types based on reason of leave taking: (1) own sickness leave, (2) maternal disability, (3) new child bonding, (4) care for an ill spouse, (5) care for an ill parent, and (6) care for an ill child. In all three states, the first two leave types are paid by the state temporary disability insurance program. The latter four types are paid by the state paid family leave program.

As noted, the comparisons are based on the five-year period from 2012 to 2016. Both ACM and Worker PLUS models are run with parameters selected to reflect each state's program rules during the period of 2012–2016, as reported in a Washington, DC paid leave economic impact report (DC Council, 2016).[vii] In addition to adjusting for these state-specific rules, the Worker PLUS model also adopts the same parameters and uses the same input data as the ACM model in its default setting to assure the results from the two models are comparable. The same input datasets are used for both models, including the 2012 DOL FMLA Employee Survey public data, the 2012–2016 ACS PUMS, and the 2014 March supplement of the Current Population Survey. Both models employ the same simulation method, logistic regression without regularization, which is the only simulation method available in the ACM model.[viii] The actual benefit outlays, program participant counts, and participant leave lengths are obtained from state program annual reports, and annual averages are derived based on 2012–2016 data for California and New Jersey and 2014–2016 data for Rhode Island, where 2014 was the first full program year.[ix] The states do not publish data on the total number of eligible workers. We therefore use estimates from an external source (DC Council, 2016) as the benchmark.

The list of parameters used in both the ACM and Worker PLUS models is included in Appendix A, while the actual state program eligibility rules are reported in Appendix B. A step-by-step guide to replicating the results from this issue brief is provided in Appendix C.

RESULTS

This section discusses the simulation results from the Worker PLUS model and ACM model as well as their comparison to the historical program data from California, New Jersey, and Rhode Island. All model statistics are reported with the sampling standard error derived from the Census Bureau ACS replication weighting procedure.[x]

TOTAL PROGRAM BENEFIT OUTLAYS

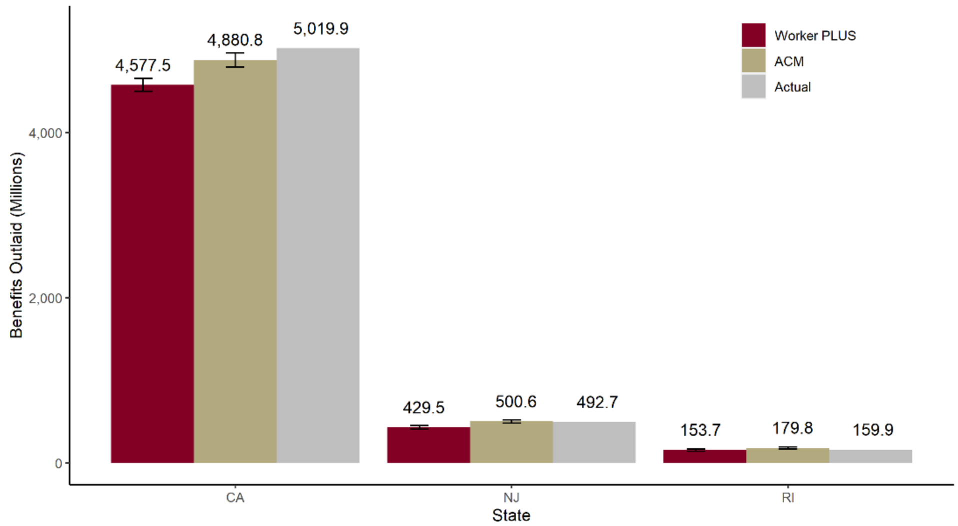
Exhibit 1 compares each model's simulated annual benefit outlays with actual annual average state-reported outlays. The comparisons suggest mixed relative performance of the two models. For California, New Jersey, and Rhode Island, the estimated outlays from the ACM model are $4,881 million, $501 million, and $180 million, respectively. The deviations from the actual outlays are respectively -2.8%, +1.7%, and +12.6%. The estimated outlays from the Worker PLUS model are respectively $4,578 million, $430 million, and $154 million. The deviations are respectively -8.8%, -12.8%, and -3.8%. These results show that both models, with a complex simulation engine built-in, can produce benefit outlay estimates that are similar to actual program data, with a maximum deviation within 13% across all comparisons.

In **Exhibit 1**, all outlay estimates are produced by the traditional logistic regression, the common simulation method chosen to benchmark the two models and the only simulation method available from the ACM model. With this comparison alone, there is no clear evidence that either model outperforms the other in predicting benefit outlays. We therefore supplement this comparison by showing outlay estimates produced by a suite of machine learning-based simulation methods, a unique feature that is only available from the Worker PLUS model. These alternative estimates are reported in **Exhibit 2**, where the estimates in the row labeled *Logistic Regression GLM* represent those from the traditional logistic regression as plotted in **Exhibit 1**.

Estimates in **Exhibit 2** show that, for all three states, the outlay estimates obtained from multiple simulation methods form lower and upper bounds that capture the actual program outlay. For Rhode Island, larger magnitude of overestimation (over 20%) is found in six out of the eight simulation methods. To further investigate, we note that the benchmarking for Rhode Island is based on partial alignment of years between the program implementation period and the ACS sample period. Because this program began in 2014, actual program data are available only from 2014 to 2016 when we compare them with simulation results derived from ACS PUMS 2012–2016. Therefore, one reason for the overestimation of outlay in Rhode Island may be that workers were unaware of the program during its first few years.

Overall, **Exhibit 2** suggests that Worker PLUS can be a more informative model than ACM by offering estimates from multiple simulation methods, which can form bounds for projecting the actual outlay. We note that not all simulation methods are equally informative in forming these bounds. For example, the magnitude of overestimation is substantially larger for Support Vector Classifier (SVC) than other simulation methods across all three states, suggesting that SVC may not be the best candidate for forming the upper bound. While these larger deviations occur for some combinations of states and simulation methods, **Exhibit 2** also suggests that the alternative simulation methods generally lead to outlay estimates that are larger than the estimate produced by traditional logistic regression, helping model users and policy makers avoid underestimating outlays during the evaluation of the financial feasibility of paid leave programs.

Exhibit 1: Simulated Worker PLUS and ACM vs. Actual Average Annual Benefit Outlays, Traditional Logistic Regression



Note: Unweighted sample sizes are respectively 702,144, 125,616, and 17,229 for eligible workers in California, New Jersey, and Rhode Island, based on the 2012–2016 American Community Survey Public Use Microdata Sample. Each individual's state is determined by where he or she works.

Exhibit 2: Simulated vs. Actual Benefit Outlays, All Simulation Methods in Worker PLUS

	California		New Jersey		Rhode Island	
	Amount	% Difference from Actual	Amount	% Difference from Actual	Amount	% Difference from Actual
Actual Program Benefit Outlays (2012–2016 average)	$5,019.9	-	$492.7	-	$159.9	-
Logistic Regression GLM	4,577.5	-8.8%	429.5	-12.8%	153.7	-3.8%
Logistic Regression Regularized	5,760.4	14.8%	470.1	-4.6%	201.5	26.1%
K Nearest Neighbor (KNN)	5,958.3	18.7%	492.1	-0.1%	197.1	23.3%
Naïve Bayes	5,904.6	17.6%	510.0	3.5%	200.3	25.3%
Random Forest	5,997.5	19.5%	492.0	-0.1%	184.5	15.5%
XGBoost (XGB)	6,086.3	21.2%	531.5	7.9%	220.4	37.9%
Ridge	6,037.7	20.3%	504.9	2.5%	210.4	31.7%
Support Vector Classifier (SVC)	7,050.1	40.4%	597.3	21.2%	240.7	50.6%
Average of All Simulation Methods	5,921.5	18.0%	503.4	2.2%	201.1	25.8%

Note: Estimates of program outlays are produced by the Worker PLUS model R engine, using 2012 FMLA Employee Survey data and 2012–2016 ACS PUMS state samples for California, New Jersey, and Rhode Island. Actual outlay data are obtained from state program annual reports.[xi] Actual outlay for Rhode Island is only available from 2014 to 2016, as the Rhode Island program began in 2014. All outlays are in 2012 million dollars. Logistic Regression GLM represents the traditional logistic regression from the generalized linear model family, which is the only common simulation method available from both the Worker PLUS model and the ACM model.

POPULATION STATISTICS — TOTAL NUMBER OF ELIGIBLE WORKERS

Exhibit 3 compares estimates of numbers of workers who are eligible for each state's leave program. The estimates come from the ACM model, the Worker PLUS model, and an external source (DC Council, 2016). Estimation of workers eligible for a state's leave program does not require any simulation of leave-taking behavior or leave needs, because eligibility in all three states can be determined solely through original ACS variables for class of employment (such as government employee status or self-employment status) and earnings. Therefore, ACM and Worker PLUS models have identical estimates for eligible workers.

Both models underestimated the number of eligible workers in California compared to the benchmark: 15.3 million versus 17.3 million. This difference of 2 million workers consists of:

- Self-employed workers (0.5 million) that are considered ineligible by both models, and
- Workers who do not meet the annual earning requirement (1.5 million, as estimated based on ACS wage data) and thus are excluded by both models.

In practice, however, some people in these two worker groups can still be eligible for the California program because

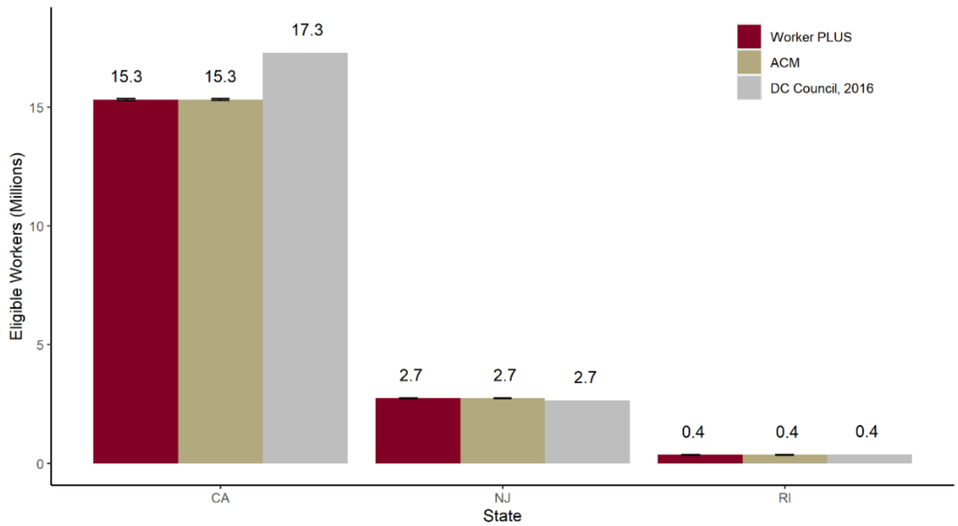
- A self-employed worker can opt into the state paid leave program through the state Disability Insurance Elective Program, and
- A worker with insufficient earnings as indicated by ACS may have sufficient earnings to be eligible based on the program's base period, a 12-month window that can be up to 6 months earlier than the ACS recall period.[xii]

Therefore, the true number of eligible workers in California should be between 15.3 million (with the two groups completely excluded) and 17.3 million (with the two groups completely included). *Without further information to determine the eligibility of these workers, excluding them is a conservative choice to avoid underestimating program outlay,* because both worker groups earn less on average than the others (e.g., self-employed workers have mean annual earnings of $28,958 compared to $52,785 for non-self-employed workers).

To account for the discrepancy in eligible worker populations between the two simulation models and the benchmark, the take-up rates for California have been adjusted upward by 13% (e.g., a take-up rate of 0.01 would be adjusted to 0.0113), the relative difference between 15.3 million and 17.3 million. This adjustment ensures that discrepancies in all other subsequent analyses would not be attributable to different assumptions made for estimating eligible worker count in California shown in **Exhibit 3**.

Overall, the two models can closely simulate the actual number of eligible workers in New Jersey and Rhode Island. However, for New Jersey, the displayed actual eligible worker count of 2.7 million in **Exhibit 3** is for eligibility for medical leave coverage (own illness and maternal disability leave), which covers fewer eligible workers due to an opt-out option for employers with private insurance. Eligibility for family leave (leave taken for new child bonding or to care for an ill child, ill spouse, or ill parent) is 45% higher at 3.8 million as estimated in the DC Council report. Therefore, for the family leave types, both models underestimate the eligible worker counts. One possibility for the underestimation is that New Jersey workers can be eligible for the program by *either* meeting the annual earning requirement *or* meeting the weekly earning requirement for at least 20 weeks, and the ACS only contains annual earning data, thus can only identify the former group of eligible workers. In our analysis, both ACM and Worker PLUS have adopted leave type-specific take-up rates for the New Jersey program to adjust for differing levels of eligibility across leave types, and so discrepancies in all other exhibits would not be attributable to different assumptions made for estimating the eligible worker count in New Jersey.

Exhibit 3: Estimates of Eligible Worker Counts from Models vs. Estimates from an External Source



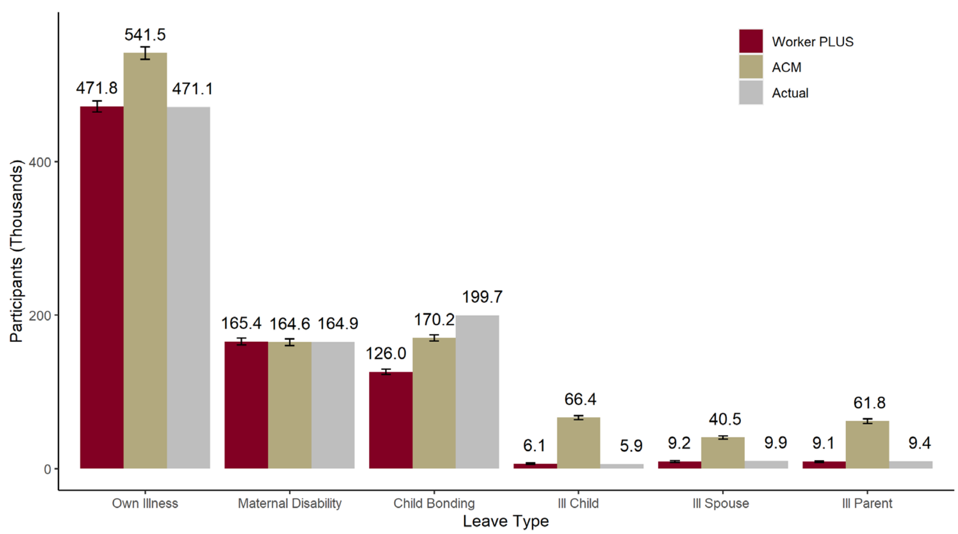
Note: Unweighted sample sizes are respectively 702,144, 125,616, and 17,229 for eligible workers in California, New Jersey, and Rhode Island, based on the 2012–2016 American Community Survey Public Use Microdata Sample. Each individual's state is determined by where he or she works.

POPULATION STATISTICS — TOTAL NUMBER OF PROGRAM PARTICIPANTS

The comparisons of program participant counts for each leave type in each state are presented in **Exhibit 4** through **Exhibit 6**. In all states, the Worker PLUS model can accurately simulate participant counts across leave types, except for the prevailing underestimation for *Child Bonding*. Again, the underestimation occurs due to the prediction of fewer leave takers by the logistic regression simulation method, which is chosen for benchmarking against the method used by the ACM model. Participant counts would be accurately predicted for all leave types under the alternative machine learning-based simulation methods available from Worker PLUS.

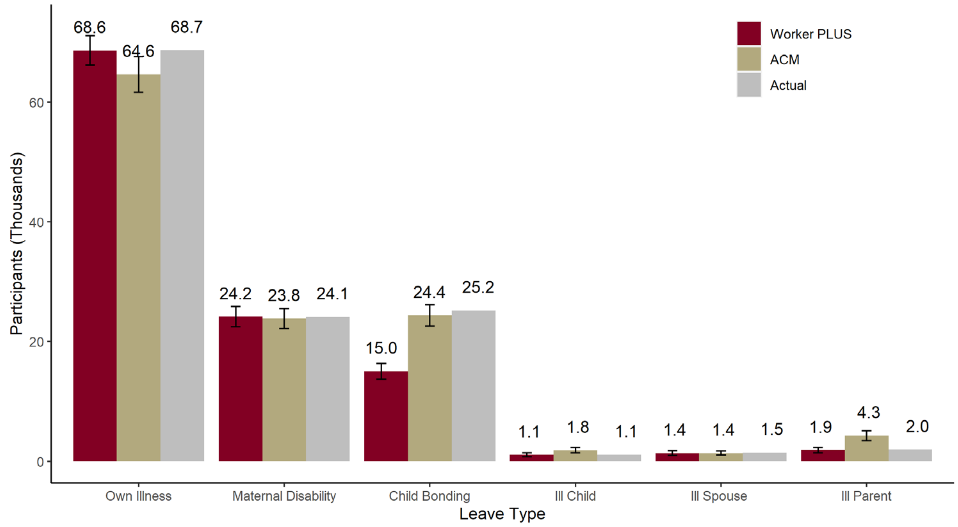
In contrast, the ACM model cannot predict participant counts as accurately. The deviations from actual counts are particularly large for the three leave types for ill family members (*Ill Child, Ill Spouse*, and *Ill Parent*). In California, the ACM model produces participant count estimates that are as large as 4 to 11 times the actual counts. These large discrepancies arise from the adoption of take-up rates in the ACM model that are not calibrated toward historical program data (which is a key process of outlay estimation in the Worker PLUS model), but are heuristically determined as a proportion of the simulated leave takers.

Exhibit 4: Simulated vs. Actual Participating Leave Taker Counts in California



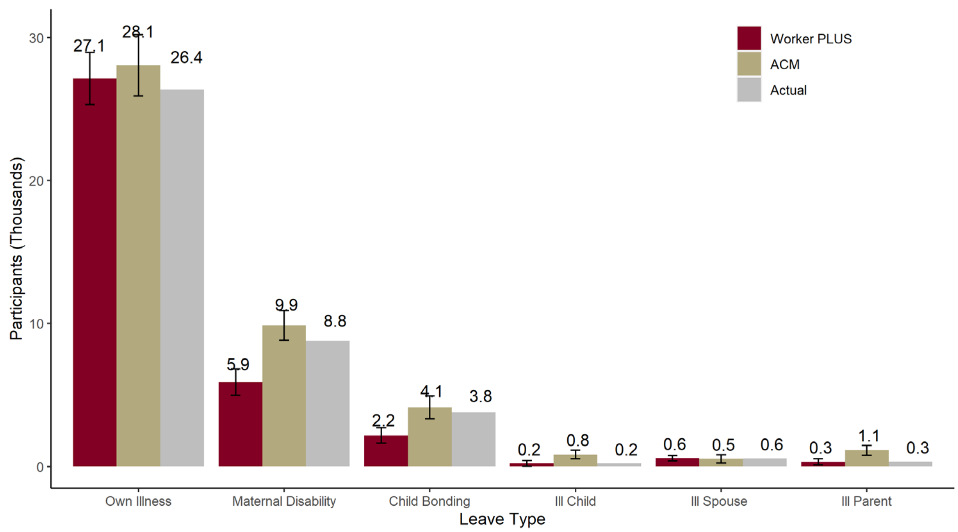
Note: Unweighted sample size is 702,144 for eligible workers in California, based on 2012–2016 American Community Survey Public Use Microdata Sample. Data are for individuals who work in California.

Exhibit 5: Simulated vs. Actual Participating Leave Taker Counts in New Jersey



Note: Unweighted sample size is 125,616 for eligible workers in New Jersey, based on 2012–2016 American Community Survey Public Use Microdata Sample. Data are for individuals who work in New Jersey.

Exhibit 6: Simulated vs. Actual Participating Leave Taker Counts in Rhode Island



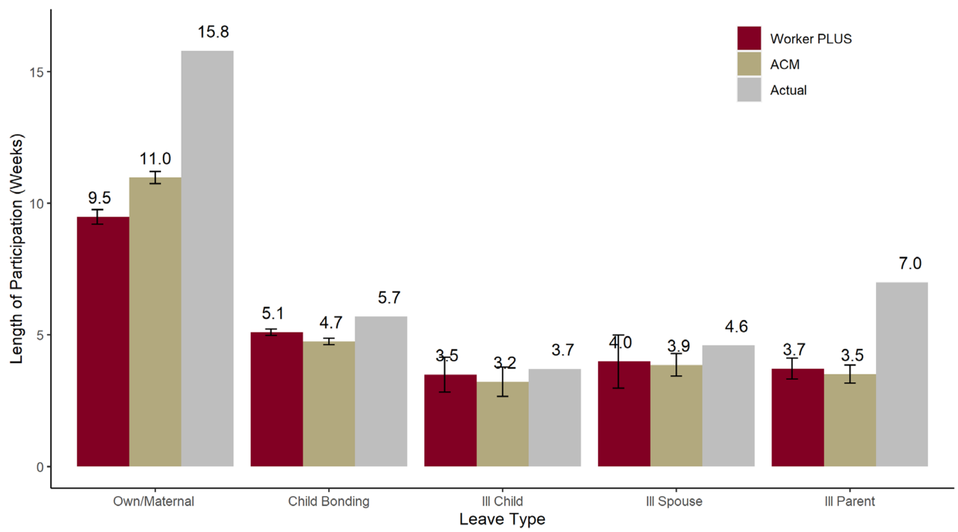
Note: Unweighted sample size is 17,229 for eligible workers in Rhode Island, based on the 2012–2016 American Community Survey Public Use Microdata Sample. Data are for individuals who work in Rhode Island.

POPULATION STATISTICS — AVERAGE LEAVE LENGTHS

Of the three states, only New Jersey reported weeks of participation by leave type. **Exhibit 7** shows the simulated versus actual mean leave lengths in New Jersey. Leave lengths for *Own Illness* and *Maternity Disability* types are not separately reported in New Jersey program annual reports, and we combine them as an aggregated category *Own/Maternal* in the exhibit. Overall, both models simulate shorter lengths for the combined leave type *Own/Maternal*. The underestimation of leave lengths is smaller in magnitude for other leave types, for which the Worker PLUS model also outperforms the ACM model.

To reconcile these differences, we note that the New Jersey data may not be directly comparable to the simulated quantities. For example, although state regulations only permit up to six weeks of paid *Ill Parent* leave, the average length of seven weeks suggests that the data contain leaves with lengths of more than six weeks. This inconsistency implies that New Jersey may have recorded leave length statistics in a different way than that reported in the program coverage rules. One possibility is that the state has double-counted leave types (e.g., an individual who takes leave for both *Ill Parent* and *Own Illness* leave has his or her leave counted for both leave types, which both ACM and Worker PLUS models do not allow). Another possibility is that the state has aggregated unpaid weeks and paid weeks from the benefit application forms. Whether these factors can fully account for the underestimation of both models is unknown without New Jersey's publishing more information on how it derives its leave lengths.

Exhibit 7: Simulated vs. Actual Mean Participation Lengths in New Jersey



Note: Actual leave length statistics by leave type are not available for California and Rhode Island. *Own/Maternal* represents an aggregated category for leaves taken due to either the worker's own illness or maternity disability, because New Jersey program annual reports do not report leave lengths separately for these leave types. Unweighted sample size is 125,616 for eligible workers in New Jersey, based on the 2012–2016 American Community Survey Public Use Microdata Sample. Data are for individuals who work in New Jersey.

CONCLUSION

This benchmarking study shows that the Worker PLUS model can predict program outlays and eligible worker counts in similar accuracy as the existing ACM model for the three state-run worker paid leave programs in California, New Jersey, and Rhode Island. The Worker PLUS model predicts program participant counts more accurately than the existing ACM model by directly calibrating take-up rates toward actual participant counts reported by states. While the logistic regression simulation method may sometimes lead to underestimation of leave takers, the options of machine learning-based simulation methods offered by the Worker PLUS model allow users to produce alternative estimates that can form lower and upper bounds of program outlay estimates. Without this flexibility in choice of simulation methods, ACM model users would risk obtaining a single version of biased estimates that may misinform policy decisions.

While the benchmarking results in this brief generally support the validity of the Worker PLUS model, model users and policy makers should be aware that outcomes simulated by the model can only approximate program implementation in the real world. Therefore, we do not recommend using the simulated numbers for accounting purposes. Instead, the model results are more suitable for feasibility analyses, early-stage planning, estimating worker population coverage, assessing the magnitude of program caseload, and evaluating the impact of a program on workers' leave-taking behaviors.[xiii]

APPENDIX A: MODEL PARAMETERS USED

ACM Model Parameters Used

Parameter	California	New Jersey	Rhode Island
ELIGIBILITYRULES	a_earnings=300	a_earnings=8400	a_earnings=3840
EXTENDLEAVES	Yes	Yes	Yes
GOVERNMENT	Yes	No	No
MAXWEEKS	OH=52, MD=52, NC=6, IC=6, IS=6, IP=6	OH=26, MD=26, NC=6, IC=6, IS=6, IP=6	OH=30, MD=30, NC=4, IC=4, IS=4, IP=4
extendproportion	OH=0.7, MD=1.0, NC=0.7, IC=0.25, IS=0.25, IP=0.25	OH=0.7, MD=1.0, NC = 0.7, IC = 0.25, IS = 0.5, IP = 0.5	OH=0.7, MD=0.7, NC=0.7, IC=0.25, IS=0.25, IP=0.25
extenddays	OH=50, MD=50, NC=30, IC=10, IS=10, IP=10	OH=40, MD=40, NC=20, IC=10, IS=20, IP=10	OH=30, MD=30, NC=15, IC=10, IS=10, IP=10
extendprob	OH=0.7, MD=1.0, NC=0.7, IC=0.25, IS=0.25, IP=0.25	OH=0.7, MD=1.0, NC=0.7, IC=0.25, IS=0.5, IP=0.5	OH=0.7, MD=0.7, NC=0.7, IC=0.25, IS=0.25, IP=0.25
topoff_min_length	20	20	20
topoff_rate	0.06	0.06	0.06
REPLACEMENTRATIO	0.55	0.66	0.6
STATEOFWORK	CA	NJ	RI
TAKEUPRATES	OH=0.40, MD=1.0, NC=1.0, IC=0.50, IS=0.85, IP =0.22	OH=0.33, MD=0.85, NC=0.85, IC=0.06, IS=0.08, IP=0.0005	OH=0.75, MD=1.0, NC=0.90, IC=0.005, IS=0.4, IP =0.005
WAITINGPERIOD	1	1	1

Note: Leave reasons are abbreviated in parameter names as follows: OH = Own Health, MD = Maternity Disability, NC = New Child, IC = Ill Child, IS = Ill Spouse, and IP = Ill Parent.

Worker PLUS Model Parameters Used

Parameter	California	New Jersey	Rhode Island
ann_hours	NULL	NULL	NULL
alpha	0.75	0	0
bene_effect	FALSE	FALSE	FALSE
base_bene_level	0.55	0.66	0.6
bond_uptake	0.0130	0.0092	0.0104
dependent_allow	0	0	c(0.07, 0.07, 0.07, 0.07, 0.07)
dual_receiver	1	1	1
Earnings	300	8400	3840
ext_base_effect	TRUE	TRUE	TRUE
ext_resp_len	TRUE	TRUE	TRUE
extend_days	0	0	0
extend_prob	0	0	0
extend_prop	0	0	0
fmla_protect	FALSE	FALSE	FALSE
full_particip	FALSE	FALSE	FALSE
GOVERNMENT	TRUE	FALSE	FALSE
illchild_uptake	0.0004	0.0004	0.0006
illparent_uptake	0.0006	0.0007	0.0009
illspouse_uptake	0.0006	0.0005	0.0016
impute_method	Logistic Regression GLM	Logistic Regression GLM	Logistic Regression GLM
matdis_uptake	0.0108	0.0088	0.0274
maxlen_bond	30	30	20
maxlen_DI	260	130	150
maxlen_illchild	30	30	20
maxlen_illparent	30	30	20
maxlen_illspouse	30	30	20
maxlen_matdis	260	130	150
maxlen_own	260	130	150
maxlen_PFL	30	30	20
maxlen_total	260	130	150
minsize	NULL	NULL	NULL
own_uptake	0.0308	0.0250	0.0823
random_seed	12312	12312	12312
rr_sensitive_leave_length	FALSE	FALSE	FALSE
SELFEMP	FALSE	FALSE	FALSE
wait_period	5	5	5

Parameter	California	New Jersey	Rhode Island
week_bene_cap	1216	594	795
week_bene_cap_prop	NULL	NULL	NULL
week_bene_min	50	0	89
weeks	NULL	NULL	NULL

APPENDIX B: PAID LEAVE ELIGIBILITY IN CALIFORNIA, NEW JERSEY, AND RHODE ISLAND

State	Eligibility Rules
California	Employed or looking for work and earned at least $300 in payroll tax wages during base period. Generally excludes self-employed, employees of religious organizations, certain domestic workers, consultants, salespeople, and students.
New Jersey	Minimum of 20 weeks with earnings of $168 or more or have earned $8,400 or more in covered New Jersey employment during the 52 weeks preceding the event. Generally excludes federal government and New Jersey county and municipal government employees.
Rhode Island	Minimum earnings of $11,520 in base period; or $1,920 in base period quarter and total base period wages of at least 1.5 times highest earning quarter, and at least $3,840 in base period. Can qualify for disability by employment and a certified disability. Generally excludes government; sole proprietors; employees of religious organizations; salespeople; certain domestic workers; and interns.

Source: Office of the Budget Director, Council of the District of Columbia (2016). Economic and Policy Impact Statement: Universal Paid Leave Amendment Act of 2016. (B21-415). Retrieved from http://lims.dccouncil.us/Download/34613/B21-0415-Economic-and-Policy-Impact-Statement-UPLAA3.pdf.

Note: For consistency with the simulation timeframe of 2012 to 2016, we applied eligibility rules derived from the cited report (i.e., current as of 2016) rather than eligibility rules as of 2020 in these states.

APPENDIX C: GUIDE TO REPLICATING RESULTS IN THIS ISSUE BRIEF USING THE WORKER PLUS R MODEL

ACM legacy modeling functions are not available in the Worker PLUS GUI, and the replication of results of this issue brief should be performed in R natively (e.g., using RStudio to run the .R files). The code required to replicate the simulations and analysis is available upon request. There are three code files: *benchmark_sim.R, benchmark_analysis.R,* and *benchmark_graphs.R.* Note that not all parameters are explicitly specified in the code; the omitted ones would be assigned the default values automatically.

Model users should follow the steps below to replicate the results in this brief. The results can be replicated using the random seed value specified below on a Windows 10 Pro operating system with OS Build version 17134.1792. The results may vary slightly for different Windows operating systems and OS Build versions.

1. In addition to the R libraries required by the Worker PLUS model (which should be installed automatically when code is run), users should ensure that the following packages are installed: *ggplot2*, *reshape2*, *varhandle*, and *stringr*.
2. Users should place the three .R files in the same directory as folders for data input. For example, *benchmark_sim.R* should be placed in the same directory as "*data,*" a folder that contains FMLA, Current Population Survey, and ACS data input files.
3. Users should place the main R model files, such as *0_master_execution_function.R* in the same directory as in Step 2.
4. Create an *exhibits* folder in the same directory as in Step 2.
5. Run the files in the following order:
   a. *benchmark_sim.R*
   b. *benchmark_analysis.R*
      i. Before running this file, the file paths in lines 14–16 should be verified so that they point to the output files generated in Step 5a.
   c. *benchmark_graphs.R*
6. After performing the above steps, all exhibits shown in this report will be generated in the *exhibits* folder.
7. To produce the numbers included in Exhibit 2, run *benchmark_sim_all_meth.R* and examine the *prog_cost* meta output file for each method and state.

Program benefit outlays are calculated from intermediate outcome variables as follows:

- Length of participation for each leave type (variables *cpl_own*, *cpl_matdis*, *cpl_bond*, *cpl_illchild*, *cpl_illspouse*, and *cpl_illparent*) is simulated.
- Wage replacement rate (*effective_rrp*) based on number of dependents (*ndep_kid*) and program base replacement rate (parameter *base_bene_level*) is calculated.
- Benefits for each leave type (variables *bene_own*, *bene_matdis*, *bene_bond*, *bene_illchild*, *bene_illspouse*, *bene_illparent*) are calculated by multiplying the weeks of participation (variables *cpl_own*, *cpl_matdis*, *cpl_bond*, *cpl_illchild*, *cpl_illspouse*, and *cpl_illparent* divided by 5 working days) for that leave type by each observation's weekly wage (*wage12* divided by 52 working weeks) and the individual's wage replacement rate (*effective_rrp*).
- Benefits across all leave types (*annual_benefit_all*) are calculated by summing the benefits received for each of the six leave types individually (variables *bene_own*, *bene_matdis*, *bene_bond*, *bene_illchild*, *bene_illspouse*, and *bene_illparent*).

- Finally, total program outlay is calculated by summing benefits across all leave types (*annual_benefit_all*) for all observations.

---

[i] Clayton-Matthews, Alan, and Randy Albelda (2017). Description of the Albelda Clayton-Matthews/IWPR 2017 Paid Family and Medical Leave Simulator Model.
[ii] For details on Worker PLUS model architecture, see IMPAQ (2021). Worker Paid Leave Usage Simulation (PLUS) Model User Manual.
[iii] Details of the Worker PLUS model user interface and output structures, runtime, and simulation comparison feature can be found in the user manual. Details of the machine learning methods can be found in IMPAQ (2021). Worker Paid Leave Usage Simulation (PLUS) Model. Issue Brief: Model Testing.
[iv] For comparison of program administrative and budget financing costs between Worker PLUS and actual costs from state programs, see IMPAQ (2021). Worker Paid Leave Usage Simulation (PLUS) Model. Issue Brief: Benchmarking Results of the Benefit Financing Module's Payroll Tax Estimates.
[v] This issue brief uses the R version of the Worker PLUS model, because the R version has replicated ACM legacy functionality programmed into it that allows for a side-by-side comparison with the ACM model.
[vi] Other key uses of the model include analyzing impacts of leave programs on workers, simulating payroll tax financing, and simulating proposed leave policies. They are respectively discussed in three separate briefs (*Estimating Impacts of Leave Policies on Low-Wage Workers*, *Benchmarking Results of the Benefit Financing Module's Payroll Tax Estimates*, and *A Guide to Perform Policy Simulation of Parental Leave for Federal Civilian Employees*) in the same issue brief series.
[vii] Office of the Budget Director, Council of the District of Columbia (2016). Economic and Policy Impact Statement: Universal Paid Leave Amendment Act of 2016. (B21-415). Retrieved from http://lims.dccouncil.us/Download/34613/B21-0415-Economic-and-Policy-Impact-Statement-UPLAA3.pdf
[viii] Logistic regression without regularization corresponds to the *Logistic Regression GLM* option in the model's GUI. GLM stands for generalized linear model.
[ix] Administrative program statistics are obtained from the following sources: Employment Development Department, State of California (2020). Disability Insurance Program Statistics. Retrieved from https://www.edd.ca.gov/about_edd/pdf/qsdi_DI_Program_Statistics.pdf; Employment Development Department, State of California (2020). Paid Family Leave Program Statistics. Retrieved from https://www.edd.ca.gov/about_edd/pdf/qspfl_PFL_Program_Statistics.pdf; New Jersey Department of Labor and Workforce Development (2017). Temporary Disability Insurance Workload in 2016 Summary Report. Retrieved from https://www.nj.gov/labor/forms_pdfs/tdi/TDI%20Report%20for%202016.pdf; New Jersey Department of Labor and Workforce Development (2017). Family Leave Insurance Workload in 2016 Summary Report. Retrieved from https://www.nj.gov/labor/forms_pdfs/tdi/FLI%20Summary%20Report%20for%202016.pdf; Rhode Island Department of Labor and Training (2014, 2015, 2016). TDI Annual Update.
[x] U.S. Census Bureau (2014). American Community Survey Design and Methodology. Chapter 12: Variance Estimation. Retrieved from https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_report_2014.pdf
[xi] Administrative program statistics including caseloads and benefit outlays are obtained from following sources: Employment Development Department, State of California (2020). Disability Insurance Program Statistics. Retrieved from https://www.edd.ca.gov/about_edd/pdf/qsdi_DI_Program_Statistics.pdf; Employment Development Department, State of California (2020). Paid Family Leave Program Statistics. Retrieved from https://www.edd.ca.gov/about_edd/pdf/qspfl_PFL_Program_Statistics.pdf; New Jersey Department of Labor and Workforce Development (2017). Temporary Disability Insurance Workload in 2016 Summary Report. Retrieved from https://www.nj.gov/labor/forms_pdfs/tdi/TDI%20Report%20for%202016.pdf; New Jersey Department of Labor and Workforce Development (2017). Family Leave Insurance Workload in 2016 Summary Report. Retrieved from https://www.nj.gov/labor/forms_pdfs/tdi/FLI%20Summary%20Report%20for%202016.pdf; Rhode Island Department of Labor and Training (2014, 2015, 2016). TDI Annual Update.
[xii] See https://edd.ca.gov/Disability/Calculating_DI_Benefit_Payment_Amounts.htm for determination of the base period in the California program.
[xiii] A detailed discussion on how different simulation methods suit different policy analysis needs is provided in IMPAQ (2021). Worker Paid Leave Usage Simulation (PLUS) Model. Issue Brief: Model Testing.