

Designing Optimal Recommended Budgeting Thresholds for a Medicaid Program

Morgan Henderson, PhD; and Ian Stockwell, PhD

In recent years, Medicaid has undergone a “rebalancing”—that is, a shift from institution-based care to home- and community-based services (HCBS).^{1,2} A key component of the rebalancing effort was the establishment of the Community First Choice (CFC) program as part of the 2010 Affordable Care Act.³ CFC is a Medicaid state plan option that allows states to offer HCBS, as opposed to institution-based services, to individuals in need of long-term services and supports.⁴ To date, uptake of the program is relatively low; as of 2018, only 8 states had enacted CFC.⁵

The relatively low uptake may be attributed to states’ concern over the cost of the program. Unlike other waiver-based programs offering HCBS, which can cap enrollment to control costs, “the statewideness and comparability requirements under the state plan are not waived” by CFC.⁴ This implies that any Medicaid beneficiary meeting the CFC eligibility requirements is eligible for CFC benefits, thus exposing the state to financial risk in terms of both enrollment and utilization.

As part of Maryland’s adoption of CFC in 2014, the state implemented a “recommended budget” system that was designed to contain costs. Individuals who meet the eligibility criteria for CFC are, based on their health status, assigned a recommended budget for their use of CFC services. These recommended budgets are not binding: The enrollee and their supports planner—that is, an individual who coordinates services in conjunction with, or on behalf of, the enrollee⁶—work together to develop a plan of services that is suitable to the needs of the enrollee. The recommended budget is intended only to facilitate the development of “tailored service plans appropriate to individual enrollees’ level of need while minimizing the likelihood of overutilization and unnecessary expenditures.”⁵ It is important to note that although plans of service can be updated often, they are not subject to influence from service providers.

Maryland’s recommended budget system was initially designed to reflect the mean CFC expenditures by acuity grouping, under the rationale that the historical mean expenditure for a particular acuity grouping is a good estimate of the services needs of a future member of that grouping and therefore serves as a useful guide for the development of a plan of service for that future individual.

ABSTRACT

OBJECTIVES: To develop and test a methodology for optimally setting automatic auditing thresholds to minimize administrative costs without encouraging overall budget growth in a state Medicaid program.

STUDY DESIGN: Two-stage optimization using administrative Maryland Medicaid plan-of-service data from fiscal year (FY) 2019.

METHODS: In the first stage, we use an unsupervised machine learning method to regroup acuity levels so that plans of service with similar spending profiles are grouped together. Then, using these regroupings, we employ numerical optimization to estimate the recommended budget levels that could minimize the number of audits across those groupings. We simulate the effects of this proposed methodology on FY 2019 plans of service and compare the resulting number of simulated audits with actual experience.

RESULTS: Using optimal regrouping and numerical optimization, this method could reduce the number of audits by 10.4% to 36.7% relative to the status quo, depending on the search space parameters. This reduction is a result of resetting recommended budget levels across acuity groupings, with no anticipated increase in the total recommended budget amount across plans of service. These reductions are driven, in general, by an increase in recommended budget level for acuity groupings with low variance in plan-of-service spending and a reduction in recommended budget level for acuity groupings with high variance in plan-of-service spending.

CONCLUSIONS: Using machine learning and optimization methods, it is possible to design recommended budget thresholds that could lead to significant reductions in administrative burden without encouraging overall cost growth.

Am J Manag Care. 2022;28(7):342-347. doi:10.37765/ajmc.2022.89180

It is possible for enrollees' actual budgets to exceed the recommended budget. When this occurs, a costly "exception" process takes place, in which supports planners must assemble documentation to justify the additional expense. Under the current system, approximately 66% of all plans of service exceed the recommended budget, thus necessitating a fuller review. In fiscal year (FY) 2019, enrollees incurred almost 15,000 approved plans of service, which suggests that the recommended budget system, initially intended to control costs, may be generating a nontrivial administrative burden.

Therefore, the objectives of this research are to determine a potential policy alternative to the current recommended budgeting methodology and to simulate the gains in administrative efficiency on actual Maryland Medicaid data from FY 2019. More broadly, our hope is that this study, although focused on Maryland, can potentially serve as a guide to other states that are considering adoption of the CFC program.

Institutional Background

Prior to the establishment of the CFC program, HCBS in Maryland were offered through 2 channels: Medicaid 1915(c) waivers and state plan options.⁷ Through these options, Maryland Medicaid offered a variety of HCBS to eligible participants, ranging from assisted living to assistive technology to personal care. Before the implementation of CFC in Maryland in 2014, almost 14,000 individuals per year accessed HCBS services.⁸

This multiplicity of programming, however, led to nontrivial administrative burden, as different programs covered similar HCBS (particularly, personal care) but reimbursed providers at different amounts; the implementation of CFC was seen as a means of streamlining and standardizing access to these services.⁵ Although 1915(c) waivers and state plan options continued to exist in Maryland following the 2014 establishment of the CFC, personal care services were largely shifted from the disparate programs to CFC. Utilization of CFC grew quickly following its implementation, rising from 4582 users in 2014 to 10,725 users in 2016.⁸

METHODS

This study uses administrative data from LTSSMaryland, an administrative database that collects enrollment, acuity, and utilization information for the universe of Medicaid HCBS beneficiaries in Maryland.⁹ All individuals in this database receive a periodic assessment that assigns Resource Utilization Groups (RUGs). RUGs are intended to be informative regarding an individual's health needs, and there are 23 RUGs that range from relatively low physical functioning impairment to relatively severe cognitive impairment.¹⁰ Although it is possible for each RUG to receive a separate recommended budget, this would entail considerable

TAKEAWAY POINTS

This study develops a novel algorithm for setting automatic auditing thresholds in a Medicaid program in Maryland.

- ▶ The Community First Choice program is a state plan option for which enrollment cannot be limited; therefore, state budgets are exposed to financial risk.
- ▶ The methodology developed in this article could enhance the "recommended budget" system by resetting these thresholds to minimize the probability of enrollees triggering automatic audits while also limiting recommended budget spending growth.
- ▶ This methodology could reduce audits by up to 36.7% by reallocating recommended budget spending, leading to nontrivial administrative efficiency gains.

administrative complexity; therefore, RUGs are currently aggregated into 7 groupings. Each of the 7 groups has an FY 2019 recommended budget, ranging from \$9075 to \$83,134.¹⁰

As noted earlier, when an individual's actual plan-of-service budget exceeds their recommended budget, a review process takes place, in which their supports planner must assemble documentation to justify the additional expense. This documentation includes a list of tasks that the personal assistant will perform and the estimated time for performance; availability of informal supports; most recent rehabilitation, hospital, or institution discharge documentation; nurse monitoring/nursing supervision notes (minimum of 2 recent); a list of other services in place; recent notes or letters from a doctor including treatment details, diagnosis, or other medical information to support the request; and services from other state or community programs.¹⁰

We approached this problem by dividing it into 2 parts. First, how can optimal RUG aggregations be created? Second, given these aggregations, how can recommended budget levels be set so that the number of exceptions is minimized without encouraging overall budget growth? Note that without this latter constraint, the problem is trivial: The solution would be for recommended budgets to be set so high that exceptions would be very unlikely to occur. The solution budget is a set of k RUG aggregations (each consisting of different RUGs) and k recommended budgets, one for each RUG aggregation. The budget will minimize the number of exceptions incurred while restricting overall recommended budget growth.

RUG Regrouping

It is possible, a priori, to recognize 2 features of the solution budget. First, RUG aggregations that include larger numbers of individuals will have more weight in the final solution than RUG aggregations with smaller numbers of individuals. Consider, at the extreme, the situation of only 2 RUG aggregations, in which the first RUG aggregation contains 1 individual and the second RUG aggregation contains all other individuals. The solution recommended budget would be focused almost entirely on minimizing exceptions in the latter group, given its disproportionate share of the population.

Second, the solution RUG budget also depends on the distribution of CFC spending within RUG aggregation. Consider another extreme example of 2 equally populous RUG aggregations, in

METHODS

which mean CFC spending is equal across groups, but the variance differs markedly. Assume that in the first RUG aggregation, all CFC spending is the same, implying a variance of 0, and in the second RUG aggregation, CFC spending differs widely across individuals in that group, implying a high variance. The optimal solution in this case would be to set the recommended budget marginally higher than mean CFC spending in group 1, thus eliminating all exceptions in this group, then to allocate the remaining recommended budget funds evenly across individuals in group 2 to set the recommended budget as high as possible.

Given that the optimal recommended budget will prioritize low-variance groups over high-variance groups and will minimize exceptions most efficiently for low-variance groups, it is desirable that RUG aggregations are created so that RUGs with similar distributions of flexible spending are aggregated together. The intuition is clear: It is important that a given recommended budget level apply to individuals who are as similar as possible in terms of their underlying acuity (and, therefore, service needs and subsequent spending).

To that end, we perform agglomerative hierarchical clustering across RUGs to identify the RUGs with similar plan-of-service flexible spending distributions. This unsupervised machine learning algorithm groups together similar observations in the data, and we deploy it on the deciles of CFC spending for each RUG from FY 2019. This algorithm yields a dendrogram, from which it is possible to determine the most similar RUG budget aggregations.

Optimal Recommended Budget Levels

Formally, the optimization problem is to select the recommended budget—that is, for k RUG aggregations, the levels $\{R_1, R_2, \dots, R_k\}$ (where R represents the recommended budget level for a given RUG aggregation)—so that the total number of exceptions is minimized while total recommended budget costs are constrained to some level. Exceptions are not known in advance: These depend on the distribution of CFC flexible spending for a given RUG aggregation. For any particular level of recommended budget for a given RUG aggregation, there is a probability that an individual will incur flexible spending over that recommended budget. This probability is $\Pr(B_{i,k} > R_k)$, where $B_{i,k}$ represents incurred flexible spending for individual i in RUG aggregation k . $\Pr(B_{i,k} > R_k)$ multiplied by the number of individuals in that particular aggregation (n_k) yields the expected number of exceptions for that particular recommended budget level for that particular aggregation. To constrain overall spending growth, we constrain the optimization problem so that the total recommended budget for the entire program is less than C , which is a constant that we select.

For k budget aggregations, the solution will solve the following problem:

$$\begin{aligned} \operatorname{argmin}_{R_1, R_2, \dots, R_k} & n_1 * \Pr(B_{i,1} > R_1) + n_2 * \Pr(B_{i,2} > R_2) + \dots + n_k * \Pr(B_{i,k} > R_k) \\ & \text{subject to } (n_1 * R_1) + \dots + (n_k * R_k) \leq C \end{aligned}$$

Given that the distributions of flexible spending for each k budget aggregation are empirically determined, there is no analytical solution for this problem. Therefore, we use numerical optimization methods to solve this problem.

Our solution algorithm uses the distribution of exceptions observed in the FY 2019 CFC plans-of-service data to estimate the number of exceptions that would occur for all possible recommended budget levels, in \$50 increments, in a search space defined by a fraction of the mean CFC flexible spending for each group (50%-150% or 0%-200%). We restrict the search space ranges to avoid “edge-case” solutions that divert recommended budgetary resources away from certain groups entirely. We simulate 100 million candidate budgets for a given RUG aggregation and deem as the solution budget that for which exceptions are minimized while satisfying the budget constraint. For the budget constraint, we use the amount of FY 2019 recommended budget spending that occurred in this period, thus allowing for no recommended spending growth and only a reallocation of existing recommended spending. We use an 80%/20% train-test sample split: We estimate the optimal recommended budget levels on 80% of the data, then test these on a remaining 20% holdout sample. We perform both analyses in Stata 15.0 (StataCorp).

Finally, we compare the demographic characteristics of the initial budget aggregations with the recommended aggregations from the 4-aggregation solution to better understand who might be affected by budget regrouping.

RESULTS

We deployed these methods on data from LTSSMaryland. This data set is composed of 14,989 approved CFC plans of service from July 1, 2018, to June 30, 2019. For the purposes of determining the optimal budget levels, the 80% training sample consists of 12,050 plans of service, and the 20% testing sample contains 2939 plans of service.

Step 1: Optimal RUG Regrouping

We first perform hierarchical agglomerative clustering to identify the RUGs that should be aggregated together into recommended budget aggregations. Results are presented in the [Figure](#).

We create our optimal RUG aggregations by reading down from the horizontal axis in the Figure. A level that intersects the dendrogram twice (at, for example, dissimilarity = 40,000) yields 2 groups, a level that intersects the dendrogram 3 times yields 3 groups, and so on. The results indicate that there are 2 primary aggregations in the data: those from BA1-SSA and those from CCO-SE3. These individual RUGs—for example, BA1, PA1, and CA1—are elements of the interRAI assessment system and indicate disparate levels of clinical and functional need. Imposing 3 aggregations would yield 1 large aggregation (BA1-SSA), 1 small aggregation (CCO-SSB), and an aggregation composed of only 1 RUG: SE3. Imposing 4 aggregations yields aggregations of BA-CA2, BB0-SSA, CCO-SSB, and SE3. [Table 1](#) presents a summary.

Step 2: Optimal Budget Levels

We then train our search algorithm on the training data set for each optimal budget aggregation consisting of 2, 3, 4, and 5 groupings presented in Table 1. We apply the resulting recommended budget aggregations to the 20% testing data set and present results in Table 2.

There were 1957 exceptions in the testing data set using the default recommended budget aggregations (of a total 2939 plans of service). Using the conservative 50% to 150% search space, we find that with 2 RUG aggregations, our recommended budgets would lead to 1860 exceptions, a reduction of 5.0% from baseline; using 3 aggregations, there would be 1859 exceptions, a 5.0% reduction from baseline; 4 aggregations would lead to 1753 exceptions, a 10.4% reduction from baseline, and 5 aggregations would lead to 1761 exceptions, a 10.0% reduction from baseline. Using a wider 0% to 200% search space, we find that 2 or 3 aggregations would lead to a 30% reduction from baseline, 4 aggregations would lead to a 36.7% reduction from baseline, and 5 aggregations would lead to a 33.9% reduction from baseline. Additionally, for all solution budgets in Table 2, the total recommended budget spending is below the recommended budget spending that would have occurred using the default values.

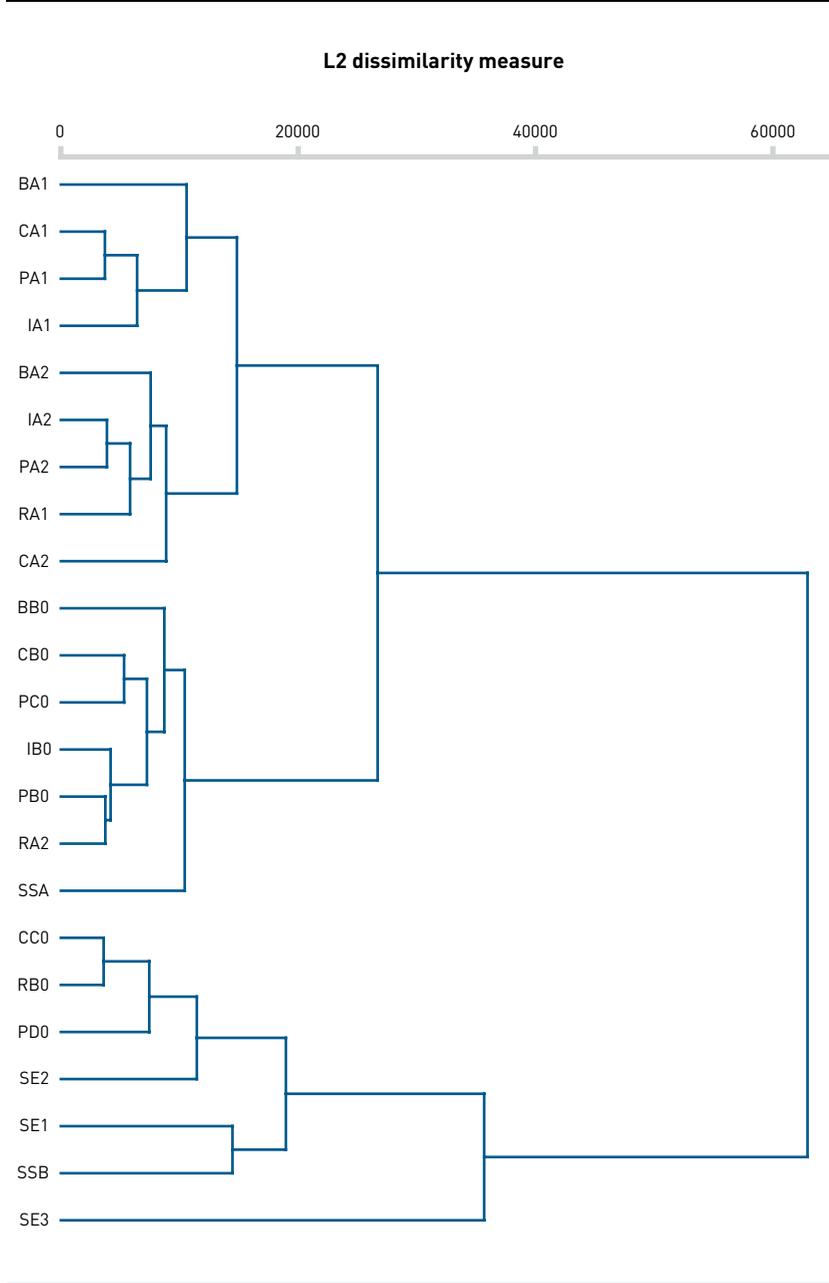
To better examine the details of the solution budget, we present the solution budgets and CFC spending from the test data set for the 4-aggregation solution using the 0% to 200% search space in Table 3.

For the first RUG aggregation, the recommended budget is \$27,300, well above the actual mean CFC spending of \$20,969. For the second RUG aggregation, the recommended budget is \$36,400—again, above the actual mean CFC spending of \$28,361. For the third RUG aggregation, however, the recommended budget is \$200 and the actual CFC mean spending is \$42,483.

Finally, we present average demographic characteristics from the initial RUG aggregation and the recommended 4-aggregation solution in Table 4.

The proposed aggregation compresses the variation across budget aggregations in demographic characteristics. Under the baseline aggregation, mean age declines almost monotonically from 70.2 years in aggregation 1 to 45.1 years in aggregation 6. Under the proposed

FIGURE. Dendrogram of CFC FY 2019 Spending by RUG



CFC, Community First Choice; FY, fiscal year; RUG, Resource Utilization Group.

TABLE 1. Optimal RUG Aggregations

Number of groups	RUG aggregation
2	BA1-SSA; CC0-SE3
3	BA1-SSA; CC0-SSB; SE3
4	BA1-CA2; BB0-SSA;CC0-SSB; SE3
5	BA1-CA2; BB0-SSA; CC0-SE2; SE1-SSB; SE3

RUG, Resource Utilization Group.

METHODS

TABLE 2. Results of Optimal Budgeting

Number of aggregations	50%-150% search space		0%-200% search space	
	Number of exceptions	% reduction from baseline	Number of exceptions	% reduction from baseline
2	1860	5.0 %	1370	30.0 %
3	1859	5.0 %	1369	30.0 %
4	1753	10.4 %	1238	36.7 %
5	1761	10.0 %	1293	33.9 %

TABLE 3. Solution Budget and CFC Spending for the 4-RUG Aggregation^a

RUG aggregation ^a	Recommended budget	Count	Actual CFC spending (mean)	Actual CFC spending (SD)
BA1, CA1, PA1, IA1, BA2, IA2, PA2, RA1, CA2	\$27,300	1198	\$20,969	\$9956
BB0, CB0, IB0, SSA, RA2, PB0, PC0	\$36,400	868	\$28,361	\$11,121
CC0, PD0, RB0, SSB, SE2, SE1	\$200	872	\$42,483	\$15,966

CFC, Community First Choice; RUG, Resource Utilization Group.

^aDue to small cell sizes, we omit aggregation 4 from this table.

TABLE 4. Demographic Characteristics by RUG Aggregation for Baseline Aggregation and Optimal 4-Aggregation Solution

Budget aggregation	Recommended budget	Mean age in years	Percent female
Baseline aggregation ^a			
1	\$9075	70.2	72.4
2	\$17,601	68.9	66.4
3	\$24,500	64.5	67.6
4	\$33,003	66.8	69.9
5	\$37,610	47.5	50.7
6	\$47,422	45.1	44.5
Optimal 4-aggregation budget solution ^a			
1	\$27,300	69.6	69.5
2	\$36,400	65.1	67.6
3	\$200	59.1	61.9

RUG, Resource Utilization Group.

^aDue to small cell sizes, we omit aggregation 7 from the baseline budget and aggregation 4 from the updated budget in this table. Age is measured as the number of years from birth date to December 31, 2019.

updated methodology, however, mean age declines from 69.6 years in aggregation 1 to 59.1 years in aggregation 3. Similarly, the gender composition of baseline aggregation declines almost monotonically, from 72.4% female for grouping 1 to 44.5% female for grouping 6. In contrast, under the optimal budget aggregation, the aggregations range from 69.5% female in grouping 1 to 61.9% female in grouping 3.

DISCUSSION

These results suggest that administrative efficiency gains are possible through reaggregating RUGs, then resetting the recommended

budgets for each RUG aggregation using numerical optimization methods. We find that recommended budgets should be set relatively high for the more populous, low-variance RUG aggregations and relatively low for the less populous, high-variance RUG aggregations. This is consistent with intuition: Spending in more populous, low-variance RUG aggregations is easier to predict, so exceptions in this aggregation are easier to prevent. Therefore, each additional dollar of recommended budget spending is used to prevent exceptions in this aggregation rather than in a high-variance aggregation.

Relative to the baseline number of exceptions, the 4-aggregation solution with a 0% to 200% search space results in a 36.7% reduction in the number of exceptions requiring automatic audits. Given that each audit requires the collection of several pieces of documentation, this reduction in the number of audits represents a potentially nontrivial gain in administrative efficiency.

These are recommended budgets rather than hard budgets. Individuals can and do exceed these budgets based on their service needs. However, to the extent that supports planners use these budgets to guide plan-of-service decisions, these altered budgets could affect true service utilization. In particular, individuals in aggregation 3 from the 4-aggregation solution (Table 3)—who would have, in the optimal solution, a recommended budget of \$200 compared with actual mean CFC spending of \$42,483—may potentially experience reduced access to services if supports planners develop plans of service that, guided by the low recommended budget, “aim low.” Additionally, this raises the possibility of greater administrative frictions for individuals in this aggregation, because, very likely, all of these individuals will

trigger the audit process. Although the extent to which supports planners use recommended budgets to guide utilization decisions is unknown, the optimization algorithm can be modified to avoid this outcome by imposing a strictly ordered solution by RUG aggregation (so that budgets with higher actual mean spending have higher recommended budgets).

Limitations

This research presents a method for setting recommended budget levels; we are unable to document, however, the extent to which the use of recommended budgets leads to realized cost savings.

Moreover, although the optimization of recommended budgets may lead to cost containment, and thus contribute to the long-term sustainability of the CFC program, we do not explicitly link this methodology to access to or quality of care.

Additionally, this methodology does not include 2 potentially relevant factors. First, we do not model the potential for strategic responses from entities in the market—for example, service providers seeking to offer additional services to participants with high recommended budgets. In this context, it is not necessary to do so: A participant's plan of services is jointly determined by the participant and the participant's supports planner without input from service providers. However, it could be the case that in other contexts, this type of induced demand may arise for individuals with high recommended budgets. Second, the model does not allow for “budget creep,” which may occur if participants and supports planners base plan-of-service decisions on the recommended budget, and instead reallocates existing recommended budget spending across individuals. This raises the potential for moral hazard: For the individuals for whom recommended budgets rise, actual service utilization may rise (regardless of health need) because of the newly “available” recommended budget. Although the extent to which this occurs is currently unknown, this suggests that an optimal recommended budget solution may be lower than that presented here to account for this phenomenon.

CONCLUSIONS

This study uses machine learning and optimization methods to develop an algorithm for setting recommended budget levels for the CFC program in Maryland Medicaid. We find that, using these methods, individuals would incur 10.4% to 36.7% fewer automatic audits, depending on the search space, thus leading to nontrivial efficiency gains without encouraging cost growth. The broader

implication is that it is possible to incorporate advanced analytics in publicly funded health care programs for the purposes of optimal program design. ■

Author Affiliations: The Hilltop Institute at UMBC (MH, IS), Baltimore, MD.

Source of Funding: Maryland Department of Health.

Author Disclosures: The authors report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

Authorship Information: Concept and design (IS); acquisition of data (IS); analysis and interpretation of data (MH, IS); drafting of the manuscript (MH); critical revision of the manuscript for important intellectual content (IS); statistical analysis (MH); administrative, technical, or logistic support (IS); and supervision (MH, IS).

Address Correspondence to: Morgan Henderson, PhD, The Hilltop Institute at UMBC, 1000 Hilltop Circle, Baltimore, MD 21250-0001. Email: mhenderson@hilltop.umbc.edu.

REFERENCES

1. Ryan J, Edwards B. Health policy brief: rebalancing Medicaid long-term services and supports. *Health Affairs*. September 17, 2015. Accessed August 21, 2020. https://www.healthaffairs.org/doi/10.1377/hpb20150917.439553/full/healthpolicybrief_144.pdf
2. Balancing long term services & supports. Medicaid.gov. Accessed August 21, 2020. <https://www.medicaid.gov/medicaid/long-term-services-supports/balancing-long-term-services-supports/index.html>
3. Burwell SM. Community First Choice: final report to Congress. Medicaid.gov. December 2015. Accessed August 21, 2020. <https://www.medicaid.gov/sites/default/files/2019-12/cfc-final-report-to-congress.pdf>
4. Community First Choice State Plan Option: technical guide. Medicaid.gov. Accessed August 27, 2021. https://www.medicaid.gov/sites/default/files/2019-12/cfc-technical-guide_0.pdf
5. Burgdorf J, Wolff J, Willink A, Woodcock C, Davis K, Stockwell I. Expanding Medicaid coverage for community-based long-term services and supports: lessons from Maryland's Community First Choice program. *J Appl Gerontol*. 2020;39(7):745-750. doi:10.1177/0733464818779942
6. Code of Maryland Regulations: 10.09.84.02.B.32. Maryland Division of State Documents. Accessed June 1, 2022. <http://www.dsd.state.md.us/comar/comarhtml/10/10.09.84.02.htm>
7. Medicaid long-term services and supports in Maryland: FY 2010 to FY 2013. The Hilltop Institute. July 28, 2015. Accessed August 25, 2021. <https://www.hilltopinstitute.org/wp-content/uploads/publications/MedicaidLTSSinMD-FY2010-2013-Vol1-ACHartBook-July2015.pdf>
8. Medicaid long-term services and supports in Maryland: FY 2012 to FY 2016. The Hilltop Institute. Revised July 8, 2019. Accessed August 25, 2021. <https://hilltopinstitute.org/wp-content/uploads/publications/MedicaidLTSSinMaryland-FY2012-FY2016-Vol5-HCBS-ChartBook-RevJuly2019.pdf>
9. Long Term Services and Supports Administration briefing on the *LTSSMaryland* tracking system. Maryland Department of Health. Accessed August 25, 2021. https://health.maryland.gov/mmcp/Documents/MMAC/2018/11_November/LTSS%20November%20Report.pdf
10. *LTSSMaryland* Database. Maryland Department of Health. Accessed August 20, 2020.

Visit ajmc.com/link/89180 to download PDF