March 2020

CENTER FOR HUMAN SERVICES RESEARCH
UNIVERSITY AT ALBANY State University of New York

Comparison Group Report

Overview

This report provides a brief overview of the comparison group developed for the New York State Medicaid Redesign Team Supportive Housing (MRT-SH) evaluation. While previous reports and analyses have focused on pre-post designs, comparing Medicaid spending and utilization before versus after MRT-SH enrollment, future reports will employ a comparison group of similar Medicaid users who are not enrolled in MRT-SH and are thus not expected to show such changes over the same interval. This document describes the data and the propensity score matching methodology used to create this Comparison group and select the final set of matched Treatment clients against whom they can be compared. Descriptive statistics are provided to assess how well these final groups are matched, and thus how these data can be most appropriately used in the upcoming analyses.

Propensity Score Matching

The Comparison group was developed using propensity score matching. Propensity score matching is a two-phase statistical technique that incorporates information regarding selection bias into estimates of an intervention's overall treatment effect.

In the first phase, the treatment selection process is examined. While various parametric and non-parametric approaches can be utilized a common approach is to use logistic regression to model treatment assignment as an outcome of covariates related to the treatment selection process. This modeling produces estimates of the likelihood of an individual receiving treatment. Information about these likelihoods or *propensity scores* are then used to match treatment and control participants to ensure "like for like" comparisons when estimating the treatment effects. To assess the usefulness of these propensity scores, post-model diagnostics are used that examine the distribution of propensity scores and compare the balance between covariates in the treatment and control groups before the data is "matched" and afterwards. If the model has achieved balance, the difference in covariates between the treatment and comparison groups should shrink after the data has been matched—ideally to the point that no statistically significant differences between the two groups remain.

In the second phase, which occurs after the modeling process has been optimized to minimize the difference in characteristics between Treatment and Comparison groups, the treatment effects are estimated. This stage will be discussed in more detail in the reports that present the results of the treatment effects analyses.

Sample and Measurement

Data Sources

Several data sources were combined to obtain MRT-SH and Random Sample client characteristics used to create the propensity scores from which final matched Treatment and Comparison groups were created (See Table 1).

Table 1. Data Sources and Variables Used in Modeling Treatment Selection

Source	Information	Notes

Medicaid Data Warehouse (MDW)	• Medicaid coverage over examined period
	• Medicaid claims
	  • Spending
	  • Diagnosis
	  • Utilization
	  • Health Home enrollment
	  • Medicaid Managed Care
	  • Medicare dual eligibility
	• County of record
	• Demographic information: date of birth, sex, race/ethnicity
	• Service date of utilization
		Claims used to determine:
		• pre-period spending (through MC_tot_rpt_plan_pd_amt, net_pd_amt_without_pgp, mcare_pd_amt, mmc_pd_amt), adjusted for inflation to 2015 dollars and capped at $1 million;
		• diagnosis (through Prim_DX_CD);
		• inpatient and emergency department utilization (through COS_CD, Rate_CD);

Child Adult Integrated Reporting System (CAIRS)	• Admission to OMH residential setting over examined period	Used as binary variable: any stay history in pre-period, versus no stay history in pre-period
Mental Health Automated Record System (MHARS)	• Admission to OMH state psychiatric facility over examined period	Used as binary variable: any stay history in pre-period, versus no stay history in pre-period
Homeless Management Information System (HMIS)	• Admission to homeless shelter over examined period	Used as binary variable: any stay history in pre-period, versus no stay history in pre-period • Regions with available HMIS data included: New York City; Long Island; Eastern New York; and the Syracuse area • Where HMIS data was not available, multiple imputation through logistic regression was used to approximate values regarding potential stays in a homeless shelter

These four datasets were deduplicated and merged, using a project-specific unique identifier. In some instances, multiple Medicaid Data Warehouse (MDW) IDs were associated with the same unique identifier. Where categorical information differed between rows (e.g., different sexes, races listed), a random MDW ID row was selected. Where possible, continuous information (e.g., cost) was aggregated across rows.

Homeless Management Information System data were not available for some regions of New York State (particularly western New York). For those regions without coverage, multiple imputation through logistic regression was used to estimate homeless shelter use.

Sample Eligibility

MRT-SH Treatment Eligibility

The original sample of MRT-SH participants used in the previous Cost and Outcomes Year 2 Volume 1 reports included 3,649 individuals in 23 programs. All participants met the original inclusion criteria of full Medicaid

coverage (no period of 60 days or longer without full coverage) for the period spanning from one year prior to program enrollment to one year after enrollment. However, six programs were determined to be inappropriate for inclusion in the Comparison group matching process and analyses, because they did not directly provide supportive housing (AIDS Institute's Supportive Services Only, HCR Capital's Access to Homes, OHIP's Senior Supportive Housing Services); focused on housing loss *prevention* rather than current lack of housing (OTDA's Eviction Prevention for Vulnerable Adults, OHIP's Senior Supportive Housing Services); or served a developmentally disabled population, for whom it would be especially difficult to identify appropriate matches from available data sources (OPWDD's Rental Subsidies Statewide, HHAP Capital's Happiness House). After these exclusions, 2,348 participants from 17 programs were retained for the Comparison group modeling.

Subsequent inclusion criteria were then applied. All clients were required to have at least one claim for a primary diagnosis of a serious mental illness (SMI), substance use disorder (SUD), HIV, or another chronic condition during their pre-period year, and to have at least some Medicaid claim cost in that year (i.e., at least some spending was required). Clients were required to have a verified county of residence in the MDW, and information on race and sex. These requirements resulted in 2,037 MRT-SH clients available for the propensity score matching process (see Table 2, below, for their distribution across MRT-SH programs).

Table 2. Distribution of Final Treatment Clients Across MRT-SH Programs

Program		N Original MRT-SH participants	N selected Treatment participants
All MRT-SH Programs		3,649	2,037
All Programs Used in Propensity Score Model		2,348	2,037
AIDS Institute	Services & Subsidies	149	117
	Pilot	17	13
HCR Capital	East 99th Street	150	107
	3361 Third Ave	34	27
	Boston Road	76	58
	Norwood Terrace	29	28
	VOA Creston Avenue	19	17
HHAP Capital	Opportunities for Broome	14	9
	Son House	26	23
	Hope Gardens	17	13
	Evergreen Loft Apartments	22	12
OTDA	Homeless Senior & Disabled Placement Program	199	146
OMH	Rental Subsidies – Brooklyn	336	290
	Rental Subsidies – Statewide	467	415
OASAS	Rental Subsidies Statewide	441	436
OHIP	Health Homes Supportive Housing Program	319	294
	Nursing Home to Independent Living (Transitions)	33	32

Comparison Group Eligibility

A random sample of 1% of New York State Medicaid users were then pulled from the MDW. However, preliminary results indicated that this random sample did not contain enough high-need, high-utilizer clients to achieve sufficient matching, particularly at the upper ends of the pre-period cost distribution. As a result, the size of the random pull was increased to approximately 5% of New York State Medicaid users.

To be eligible for inclusion in the data pull for the random sample, at least two years of continuous full Medicaid coverage between 2011 and 2017 was also required. While the MRT sample has a defined pre-period (the 12 months prior to MRT-SH enrollment), the Random Sample clients had no such external bounds because they were not enrolled in any comparable program of treatment. As such, their pre-period was defined as their first 12 month period of coverage. This resulted in a sample of 2,208,395 individuals.

Subsequent descriptive analysis suggested that this sample was still not adequate to identify an appropriate comparison group, because the MRT-SH clients are a particularly acute population. A final data pull was completed of only those Medicaid users with a pre-existing diagnosis of SMI, SUD, HIV, or a chronic condition (more than 90% of MRT clients had at least one of these diagnoses).

After the data pull, which filtered by diagnosis and continuous coverage, additional filters were applied to remove clients with no Medicaid spending during the coverage period, those without a verified county of residents listed in the MDW, and those missing valid data on race and sex. The final inclusion criteria resulted in 1,244,090 Random Sample clients available for the propensity score matching process. Table 3, below, shows the effect of each exclusion on sample size.

Table 3: Sample Size Restrictions

	MRT-SH Sample	Random Sample
Full/Original Sample	3,649	2,208,395
Program Exclusions	2,338	2,208,395
Diagnosis Exclusions	3,146	1,250,297
Medicaid spending > $0	3,524	2,129,485
Sex Exclusions	3,576	2,208,395
County Exclusions	3,570	2,198,330
Race Exclusions	3,532	1,909,973
All Exclusions	2,037	1,244,090

Propensity Score Analysis Methodology

There were several likely skews or biases anticipated in the available clients and data which had potential to affect the modeling.

First, MRT-SH program enrollment occurred on a rolling basis, with clients entering the programs between 2012 and 2016. For treatment clients, treatment year was defined as their first 12 months in the program, while for Comparison clients it was defined as their first 12 months of continuous Medicaid coverage in the 2012-2016 period (within a spell of at least two years of continuous coverage). While Medicaid costs were adjusted for inflation, other policies and interventions may have been introduced over this period, which could impact Medicaid spending for either the Treatment or Comparison groups.

Second, geography was identified as a key factor likely to influence MRT-SH enrollment. MRT-SH programs are not evenly distributed throughout New York State; thus, clients' probability of enrolling in an MRT-SH program was in part dependent upon the community in which they lived. Further, other available client services were likely to be more available in larger urban metropolises, or even in micropolitan areas, than in rural regions. Counties were thus categorized as one of five levels of urbanicity: New York City; Other Metropolitan–Central City; Other Metropolitan Non-Central-City; Micropolitan; and Rural.

Finally, there was an expectation that the Treatment and Comparison groups would have large differences in their average pre-period spending, because the MRT-SH programs were targeted explicitly to those Medicaid clients who used services intensively. To ensure sufficient overlap on the upper ends of the distribution, the identification of the Random Sample went through the iterations discussed above (i.e., initially a random sample of 1% of New York State Medicaid users was pulled from the MDW; then the random sample was increased to 5%; and finally added diagnoses as an additional sample selection criteria). Ultimately, this allowed achievement of the necessary overlap in the upper ends of the distribution.

Figure 1 illustrates the distribution of cost among the MRT-SH and Random Sample clients within decile (10%) bands of spending (see Table 5 for the bounds of these decile bands). While the shape of the distribution within each decile is not identical, there is adequate overlap between the two groups (especially within the highest four spending deciles).

Figure 1: Cost Decile Distributions
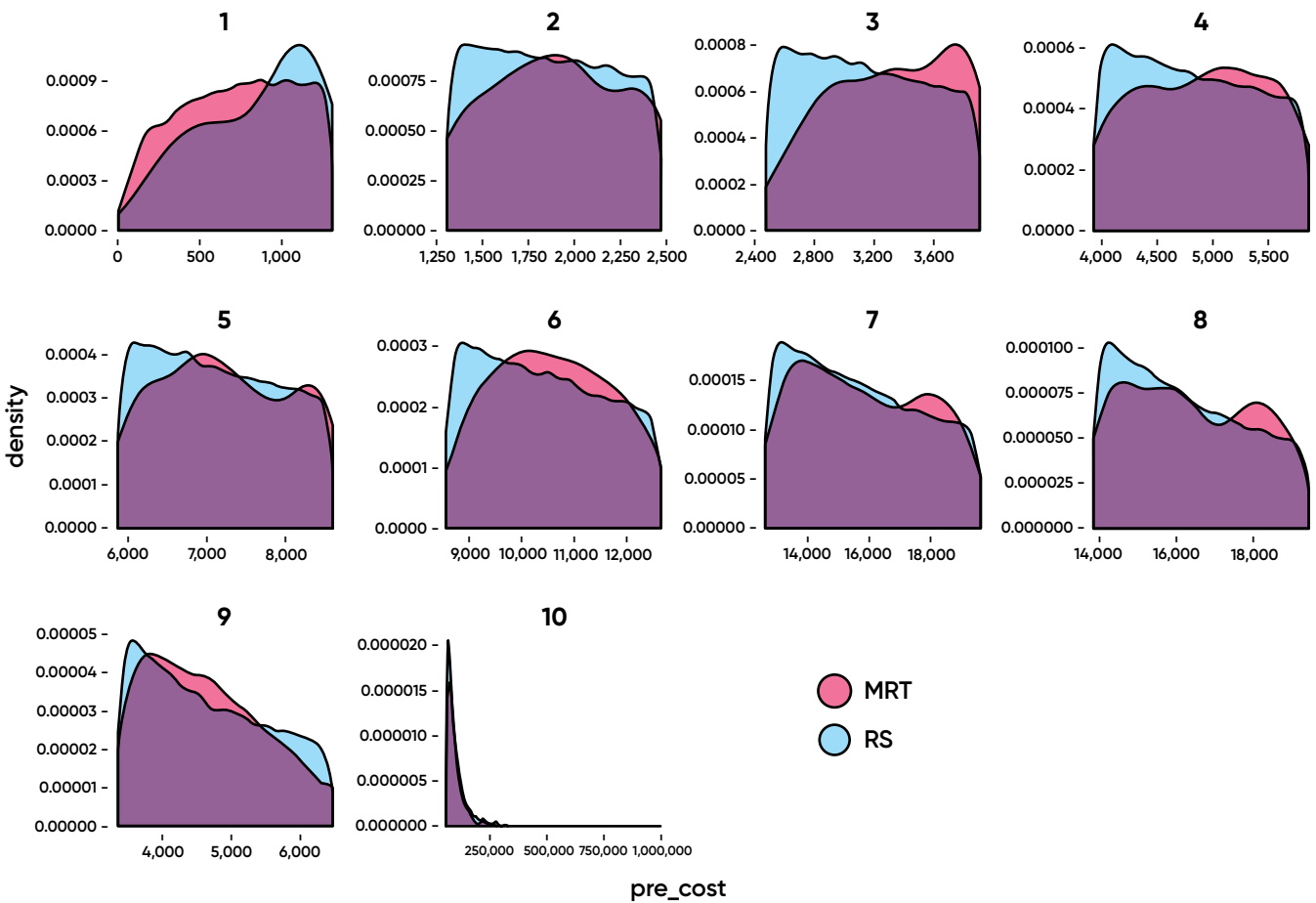
Because of their potential to bias selection effects, three variables (treatment year, urbanicity, and cost decile) were included in the model as "fixed" variables, wherein selected Treatment and Comparison pairs were required to *exactly match* on these characteristics.

Modeling Selection of Treatment Participants

Estimating Propensity Scores

This phase of propensity score analysis modeled the treatment selection process. The goal in this phase of the model was to achieve overall balance between the matched Treatment and Comparison group populations. Balance was measured by the degree to which statistically significant differences between matched pairs in the Control and Treatment groups using the logistic regression model.

All the variables listed in Table 1 were entered into a logistic regression model as predictor variables. As noted above, selected pairs were required to be in the same pre-period spending decile; have the same urbanicity; and have the same treatment year. The remaining variables (diagnoses, race, sex, age, admission to homeless shelter or OMH setting, utilization, Health Homes enrollment, managed care enrollment, and dual enrollment in Medicare,) were included as covariates where "partial" matching was allowed; for these factors, better alignment resulted in more similar propensity scores but exact matching was not required. Although we would ideally like to match clients who had the exact same profile on these characteristics, the limitations of a real-life sample mean that fixing on all variables is not possible. In practice, fixing certain variables is likely to result in greater imbalance on other variables in order to make appropriate matches.

Almost all of these variables were found to be significant risk factors for Treatment selection, meaning that the eligible MRT-SH and Random Sample demonstrated significantly different distributions pre-modeling (detailed results of the logistic regression model are included in Table A1 in the Appendix). Only pre-period admission to an OMH residential setting was not a significant risk factor for selection into MRT-SH programs.

Generated Propensity Scores

Figure 2 depicts the final distribution of propensity scores estimated for both Treatment participants and comparison group-eligible participants. The figure below shows the distribution of propensity scores after 1) the comparison eligible group was restricted to the full population of Medicaid users with a SUD, SMI, HIV, or chronic condition diagnosis and 2) covariates for health home status, Medicare use, and managed care were added to the logistic regression model. The overlap between the two groups in the graph below represents the members who are potential candidates for matching.
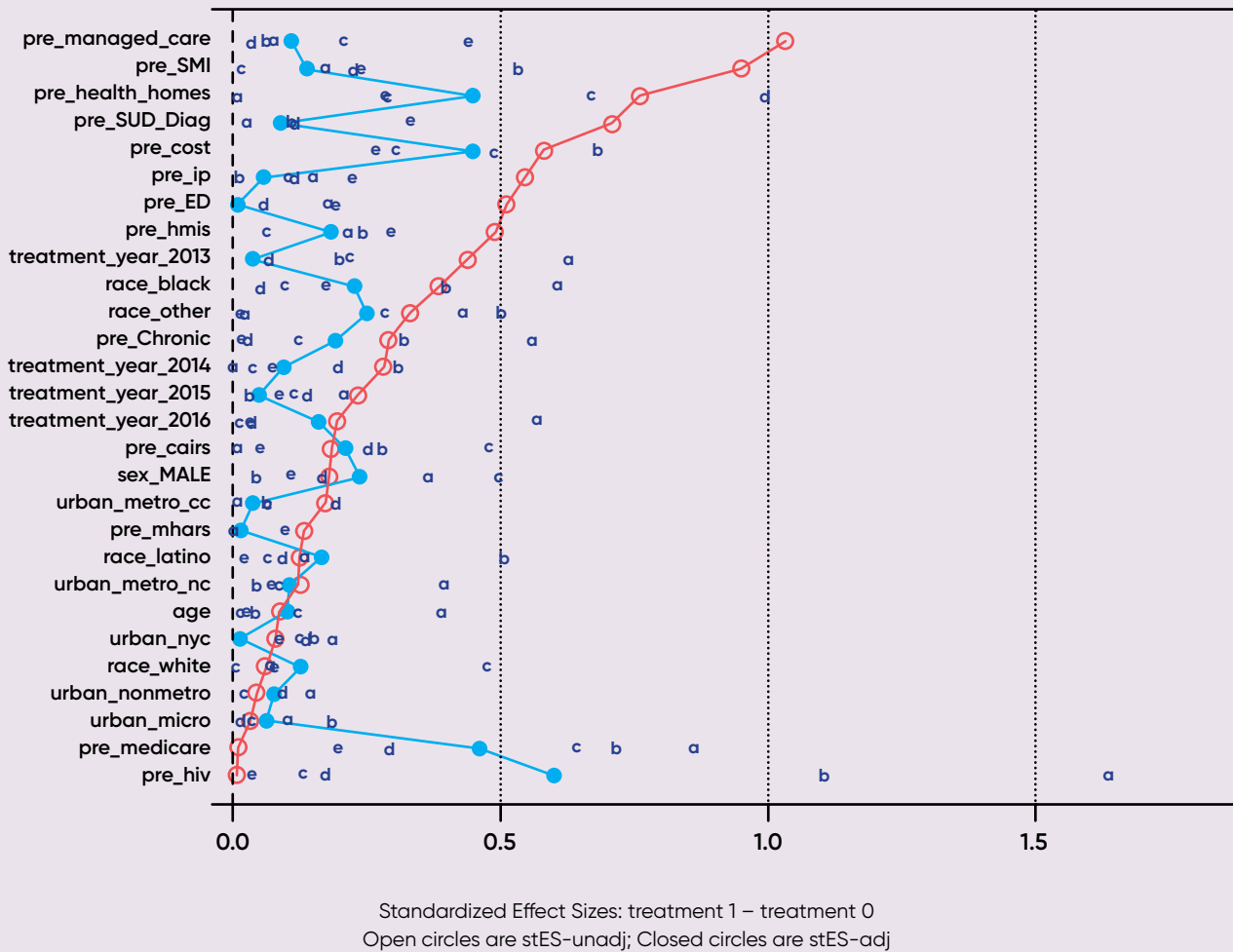
Figure 2: Propensity Score Distributions
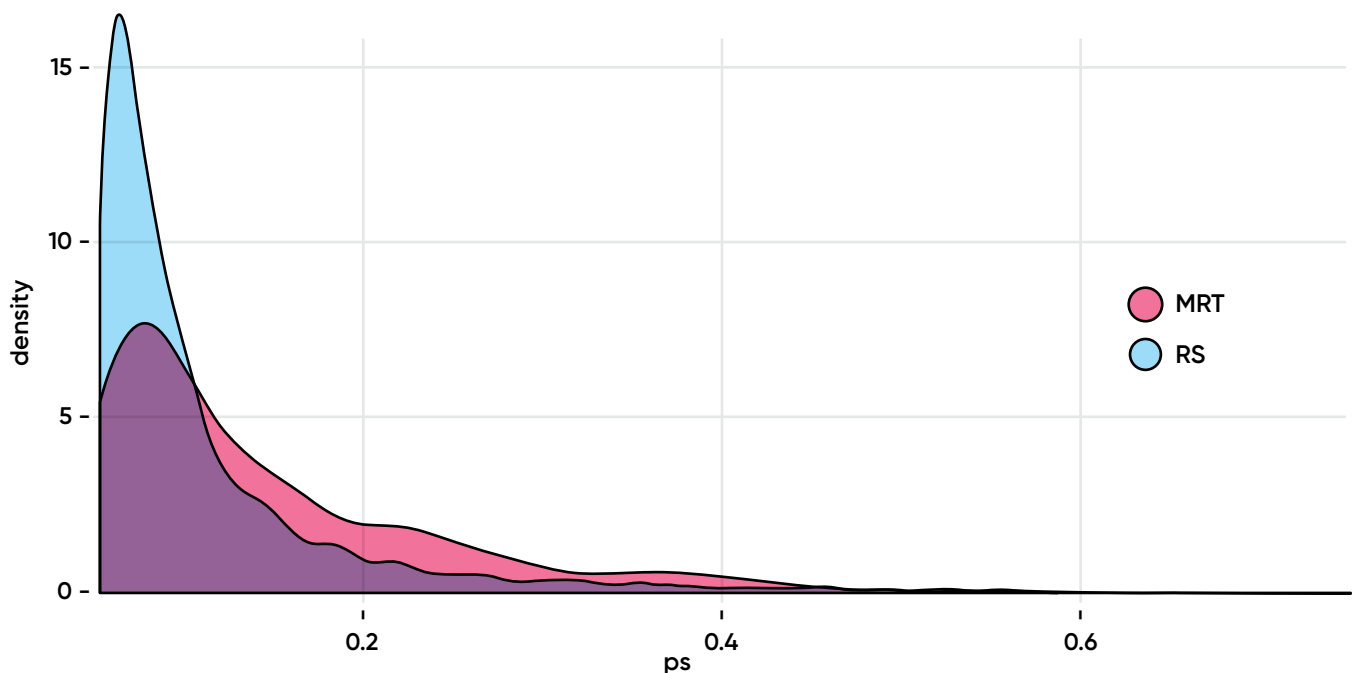
Assessing Balance

Figure 3 illustrates the improvement in balance achieved through propensity score matching. The x-axis shows the covariate effect size, while the y-axis represents each of the covariates. The red circles depict the Standardized Effect Size in the unadjusted sample, while the blue circles represent those effect sizes after adjusting for estimated propensity scores. The letters associated with each covariate represent that covariate's strata distribution. In this figure, a blue circle plotted at 0 on the x-axis for a covariate on the y-axis would represent perfect balance between control and treatment groups for that particular covariate. The gap between the blue and red circles for one covariate represents the level of change in balance achieved: the larger the gap, the larger the change in balance. For some covariates, such as age, there was very little change in balance; because age was already close to zero, there was less room for improvement in balance. For others, such as pre-period managed care, there was a large change in a positive direction; whereas some covariates, such as HIV, had a large change in a negative direction, meaning that the balance actually became worse after the matching process. While this is not ideal, it is a foreseeable consequence of fixing certain variables – when perfect matches are required for some covariates, this often requires the matches to be less "alike" on other covariates.

Overall, the plot depicts a substantial improvement in balance between treatment and comparison groups following propensity score matching. While significant improvements in balance were achieved after modeling (i.e., for most covariates, the blue circle is closer to 0 than is the red), notable imbalances remain for some covariates (i.e., the blue circle is still quite distant from 0), particularly for pre-period cost and Health Home enrollment.

Figure 3: Balance Assessment

Standardized Effect Sizes: treatment 1 – treatment 0
Open circles are stES-unadj; Closed circles are stES-adj

It should be noted that Figure 3 shows the covariates prior to matching on the fixed variables (cost decile, urbanicity, and treatment year). After these are fixed in the matching process, the difference in these three variables is reduced to zero.

Selecting Matched Pairs

Treatment and Comparison clients were selected as matched pairs (1 to 1 matching) based on a "nearest neighbor" methodology (e.g. Treatment clients were matched to the Comparison client with the closest propensity score), restricted by same cost decile, urbanicity, and treatment year. The matching methodology was carried out "with replacement", which means that while each final Treatment client has an individually-matched Comparison client, an individual Comparison client may be the match for multiple unique Treatment clients. This procedure ensures optimal matching by allowing all Treatment clients to have the best Comparison match possible, though there is the risk that it can result in over-reliance on a few Comparison clients who could then have a disproportion influence on the results.

Under these criteria, 2,037 Treatment clients were matched to 1,929 Comparison clients, resulting in 2,037 pairs or 4,074 entries. (As noted above, Comparison clients were allowed to be a match for more than one treatment client.) Subsequent diagnostic analysis was done to ensure that the matching procedure was not dependent on a few comparison group-eligible participants (as indicated by a comparison client having a large number of treatment client matches) and that the propensity scores were well-matched between pairs.

Propensity score matching results

No Comparison client was selected as a match more than four times, and typically no more than once (94.7% of Comparison clients were matched to one Treatment client, 4.7% were matched to two, 0.5% to three, and 0.1% to four), demonstrating that the process did not overly rely on any individual Comparison client.

Figure 4 shows the distribution of differences in propensity scores for matched pairs. The differences in propensity scores (Treatment – Comparison) ranged from -0.10 to 0.31, but 99.8% of the pairs had differences <0.10 in either direction. A large number of our matched pairs had a difference very close to zero (93.9% of the pairs had differences <0.01 in either direction).

Figure 4: Distribution of Differences in Propensity Scores for Matched Pairs
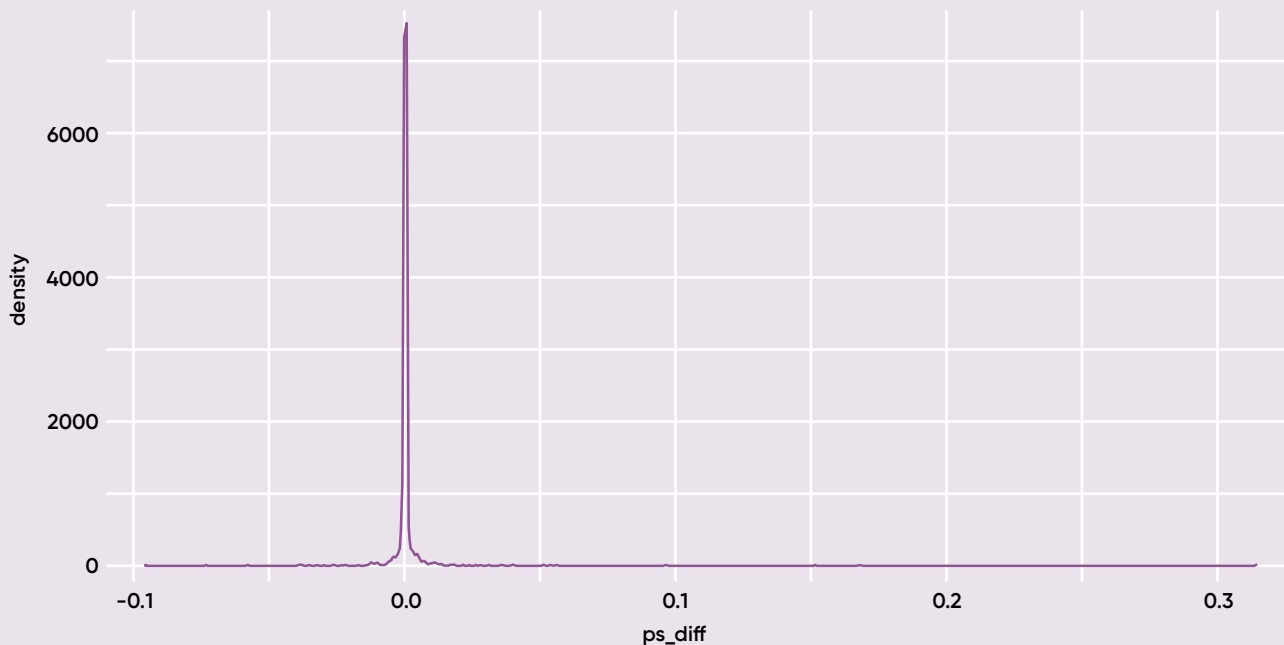
Also, after modeling and pair selection, many variables no longer significantly differed between the groups, demonstrating improved balance over the comparisons to the full sample of comparison group-eligible participants. Table 4, below, shows the characteristics of the clients in the final Treatment and Comparison groups. Clearly, these key variables are well-matched between groups, with no statistically significant differences except on pre-period inpatient use and dual Medicare eligibility (both significantly higher among the Comparison group).

Table 4. Selected Characteristics of Final Treatment and Comparison Clients (Post-Matching)

	Treatment	Comparison	Sig.
Diagnostic categories			
Serious Mental Illness	77%	78.5%	n.s.
Substance Use Disorder	51.5%	54%	n.s.
Other Chronic Condition	50%	52%	n.s.
HIV	5%	5%	n.s.
3 or more of the above	20%	21%	n.s.
All 4 of the above	1%	1.5%	n.s.
Treatment duration			
Less than 6 months	11%	--	--
Between 6 and 12 months	16%	--	--
Between 13 and 18 months	14%	--	--
Between 19 and 24 months	20%	--	--
More than 24 months	39%	--	--
Care coordination			
Medicaid Managed Care Enrollment	41%	43%	n.s.
Health Homes Enrollment	38%	38%	n.s.
Dual Eligibility	9%	12%	*
Utilization			
Any pre-period inpatient	51%	61%	***
Any pre-period ED	68%	67%	n.s.

Due to fixing the matches by cost deciles as described above, the final sample skewed heavily towards higher spenders (Table 5). This is consistent with the fact that the Treatment group is a high-spending population on average and required high-spending matches in the Comparison group.

Table 5. Spending Decile distribution for selected Treatment and Comparison clients.

Decile	Pre-Period Spending Range	N final participants in decile (Treatment + Comparison)
1	Minimum ($20) – $1,300	50
2	$1,301 – $2,474	86
3	$2,475 – $3,915	114
4	$3,916 – $5,865	232
5	$5,866 – $8,600	287
6	$8,601 – $12,605	356
7	$12,606 – $19,600	590
8	$19,601 – $33,600	827
9	$33,601 – $64,600	1025
10	$64,601 – Max ($660,500)	517
All Deciles		4,074

Conclusions

This report details selection of the Treatment clients and the creation of the matched Comparison group for use in upcoming MRT-SH Evaluation reports. A propensity score matching approach was applied, wherein appropriate matches were determined for each eligible MRT-SH client based on demographic, clinical, and utilization criteria from a random sample of New York State Medicaid users who met certain coverage, diagnosis, and spending criteria. While the original MRT-SH and Random Sample groups were quite different on most of the modeled criteria, the final Treatment and Comparison clients were not significantly different in most of their key characteristics, indicating improvements in sample balance after modeling. However, dual eligibility and inpatient utilization remained significantly different between groups even after modeling, demonstrating that while this process was able to select Comparison clients who were generally similar to the Treatment population, this Treatment group still represents a very particular, acute, and complex clientele for whom appropriate matches are likely limited. Because MRT-SH aims to target, enroll, and serve a particularly acute and complex client population, this uniqueness is not unexpected.

This model was optimized for matching by overall Medicaid spending, as this was considered the most potentially important outcome variable. As such, this appears to be a robust model for comparing differences in overall spending. Use of the matched groups to assess other outcomes, or to make comparisons using subgroups of the matched sample (e.g., within diagnostic groups) should be considered exploratory. This is because the fixing by cost decile will artificially increase imbalance on other variables (this is most evident with pre-period inpatient utilization as shown in Table 4). For this reason, future research should confirm those findings with additional models optimized for those specific comparisons.

Appendix

Table A1, below, shows the results of the logistic regression analysis used to predict treatment selection and thus estimate propensity scores.

Table A1: Summary of Logistic Regression Analysis of Predictors of Treatment Selection

	Coeff.	S.E.
Pre-period Medicaid costs (logged)	**0.468***	**0.022**
Race/ethnicity		
Non-Hispanic white (ref. category)	--	--
Non-Hispanic black (1=yes, 0=no)	0.866***	0.059
Hispanic/Latino of any race (1=yes, 0=no)	0.627***	0.072
Non-Hispanic, other race (1=yes, 0=no)	0.098	0.11
Male (1=yes, 0=no)	**-0.007**	**0.049**
Age in years	**0.008***	**0.002**
Pre-period diagnosis (any claim)		
HIV (1=yes, 0=no)	-0.415***	0.109
Severe mental illness (1=yes, 0=no)	1.491***	0.058
Substance use disorder (1=yes, 0=no)	0.987***	0.055
Selected chronic medical conditions (1=yes, 0=no)	-0.011	0.053
Pre-period housing status		
Admission to OMH residential setting (1=yes, 0=no)	-0.540***	0.134
Admission to homeless shelter (1=yes, 0=no)	0.631***	0.068
Admission to OMH acute psychiatric setting (1=yes, 0=no)	0.123	0.225
Pre-period utilization		
Any inpatient utilization (1=yes, 0=no)	-0.304***	0.056
Any emergency department utilization (1=yes, 0=no)	0.545***	0.053
Care management		
Any managed care enrollment (1=yes, 0=no)	-2.603***	0.051
Any Medicare dual enrollment (1=yes, 0=no)	-1.953***	0.084
Any Health Home enrollment (1=yes, 0=no)	1.100***	0.052
Geography		
New York City (ref. category)	--	--
Other metropolitan – central city (1=yes, 0=no)	0.371***	0.058
Other metropolitan – not central city (1=yes, 0=no)	0.002	0.083
Micropolitan area (1=yes, 0=no)	0.457***	0.112
Rural (1=yes, 0=no)	0.005	0.201
Year		
2013 (ref. category)	--	--
2014 (1=yes, 0=no)	2.093***	0.138
2015 (1=yes, 0=no)	1.559***	0.143
2016 (1=yes, 0=no)	2.039***	0.139
Constant	**-12.178***	**0.253**
Observations	**495,157**	
Log Likelihood	**-9,402.20**	
Akaike Information Criterion	**18,856.40**	

Comparison Group Report

March 2020

CENTER FOR HUMAN SERVICES RESEARCH
UNIVERSITY AT ALBANY State University of New York