Allegheny Housing Assessment: Updated Methodology Report

Prepared by the Allegheny County Department of Human Services

JANUARY 2026

TABLE OF CONTENTS

EXECUTIVE SUMMARY

In August 2020, Allegheny County Department of Human Services (DHS) implemented the Allegheny Housing Assessment (AHA), developed by Dr. Rhema Vaithianathan and partners at the Centre for Social Data Analytics (read more here). The AHA is a decision support tool that helps DHS prioritize admissions to Rapid Rehousing (RRH) and Permanent Supportive Housing (PSH). In the fall of 2024, DHS began the process of updating the model (previous updates to the model were made in November 2020). This update aims to improve model performance by expanding the training dataset and adding additional outcome metrics to improve targeting of housing resources. The three most substantial changes made to AHA as part of this update include:

- **An expanded training dataset.** The model was last trained using data from January 2017 through September 2019. This update incorporates training data from 2016 through 2024.
- **The use of inverse propensity score weighting.** To enable the use of data collected after the date of AHA implementation without biasing the model, we trained only on applications from people who were not assigned to PSH or RRH. Using weights on the data that are inversely proportional to the individual's likelihood of assignment allows the model to perform well on future scoring cohorts of people who will be assigned to housing while still producing probabilities of the outcomes without assignment.
- **The addition of a homelessness model component.** Building on evidence that RRH and PSH effectively eliminate the risk of future homelessness, the updated model incorporates risk of future homelessness into its score calculation. This new component will ensure that people with high AHA scores are among those most likely to experience homelessness and thus most likely to be helped by these supportive housing programs.

AHA produces a 10-point risk score based on several component predictive models. The outcomes that are used to create this composite risk score in the updated model are described in **Table 1**. Because the training cohort is a particularly housing-insecure group, one-year homelessness was the most common outcome for this group, at about 32%.

TABLE 1: **Descriptions of component model outcomes**

OUTCOME	DESCRIPTION	PREVALENCE
Mental Health Inpatient	At least one inpatient mental health service funded by Medicaid in the 12 months following the assessment	12.1%
Emergency Room 4+ Visits	Four or more emergency room visits in the 12 months following the assessment	23.9%
Jail Booking	At least one Allegheny County Jail booking in the 12 months following the assessment	15.6%
Any Homelessness	Any shelter usage or contact with street outreach in the 12 months following the assessment	31.6%

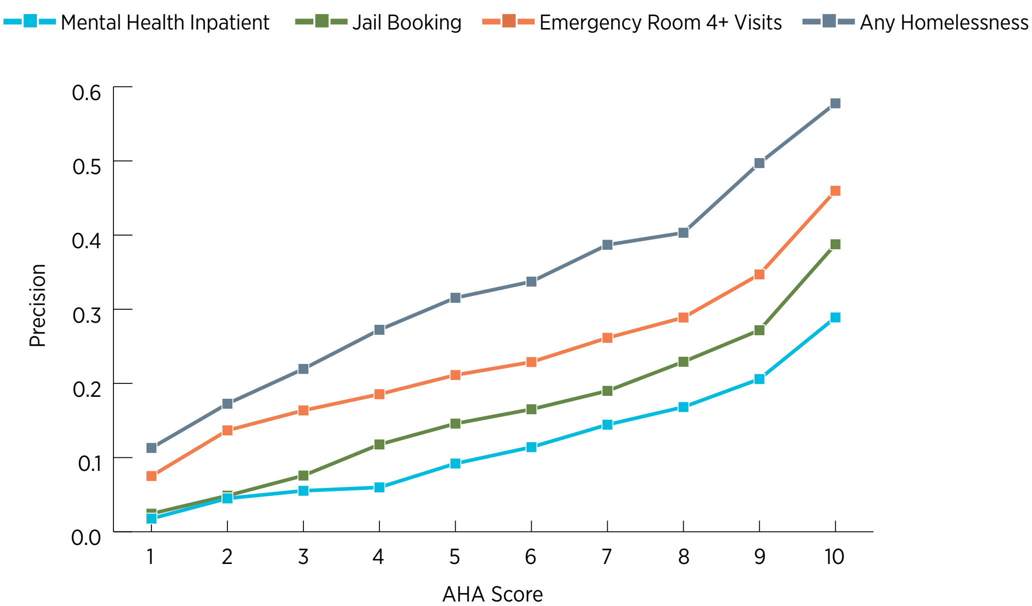
Table 2 shows the Area under the Curve (AUC) of the updated and current versions of AHA. The original three outcomes had comparable AUCs across models. The AUC for the homelessness outcome was substantially higher in the updated model, reflecting the fact that AHA was not effectively prioritizing people who were at highest risk of homelessness before this update.

TABLE 2: **AUCs of updated and current AHA, by outcome**

OUTCOME	AUC OF UPDATED AHA	AUC OF CURRENT AHA
Mental Health Inpatient	0.71	0.66
Emergency Room 4+ Visits	0.66	0.68
Jail Booking	0.72	0.73
Any Homelessness	0.68	0.54

The performance of the updated AHA model is summarized in **Figure 1**. The likelihood of each outcome increased with AHA score, demonstrating that the model effectively identified the most at-risk individuals.

FIGURE 1: **Rate of each component outcome by AHA score**



This report details the updates to the AHA's data and training methodology. It also presents updated performance metrics and robustness checks.

UPDATE DETAILS

Data

The updated training set included all RRH and PSH assessments from January 2016 through June 2024. We restricted our training exclusively to heads of household, with about 21% of clients having multiple assessments. We filtered out "soft assessments" where applicants were not assigned a Vulnerability Index-Service Prioritization Assistance Tool (VI-SPDAT) score and ensured that no individual appeared more than once in a 180-day period.

There were 18,008 rows in total in the updated training set, with 13,942 unique clients, compared to 9,213 rows in the November 2020 training data. The Jail Booking and Any Homelessness models were trained on all 18,008 rows, while the Mental Health Inpatient and Emergency Room 4+ Visits models were trained only on clients enrolled in Medicaid during the year of their assessment because those outcomes would not be observable for people who are not enrolled in Medicaid. This limits the training data for these component models to 13,673 total rows with 10,440 unique clients.

Table 3 presents a demographic profile of the training data. Note that many individuals appeared in the training data several times each. About 53% of clients assessed from January 2016 through June 2024 were Black, compared to 41% who were White. Like the current AHA training cohort, 50% of those assessed were women and 50% were men. Overall, the demographics of the Medicaid-enrolled cohort were very similar to those of the full cohort, which alleviated concerns about training component models on cohorts that were substantially different from one another.

TABLE 3: **Demographic profile of the training cohort**

		ALL 18,008	MEDICAID-ENROLLED 13,673
Race	Black	9,540 (53%)	7,815 (57%)
	White	7,285 (41%)	5,722 (42%)
	Other	613 (3%)	99 (0.7%)
	Missing	570 (3%)	37 (0.3%)
Gender	Female	8,746 (50%)	6,991 (51%)
	Male	8,820 (50%)	6,679 (49%)
	Missing	14 (0.1%)	3 (0.0%)
Household type	Single	13,140 (73%)	9,835 (72%)
	Youth	764 (4%)	541 (4%)
	Family	4,104 (23%)	3,297 (24%)
Disability (self-reported)	Yes	14,774 (82%)	11,520 (84%)
	No	3,234 (18%)	2153 (16%)

Features used to train the model came from DHS's Data Warehouse that includes integrated person and service data from a wide variety of sources. A summary of the features used to train the model is shown in **Table 4**. Not every available feature was used to train each component model. For a summary of the number of features from each category that were included in each specific subcomponent, refer to **Table 5**.

TABLE 4: **Overview of candidate model features**

CATEGORY	DESCRIPTION	NUMBER OF FEATURES CONSIDERED
Housing and homelessness	• Past shelter stays, street outreach and homelessness prevention services • Rapid Rehousing or Permanent Supportive Housing	51
Criminal/legal	• Counts of felony, misdemeanor and summary charges by type of crime (e.g., drug related, violent, property, etc.) • Number of days detained in jail	246
Behavioral health	• Inpatient or outpatient behavioral or mental health episodes • Inpatient or outpatient substance use disorder treatment • Behavioral health diagnoses	316
Physical health	• Inpatient or outpatient physical health claims • Physical treatment for substance use disorder	54
Child welfare	• Past child welfare referral(s) — accepted or screened-out • Alleged perpetrator or involved adult	72
Benefits	• Medicaid, SSI and TANF enrollment	27
Demographics	• Age, sex, race, ethnicity, education and year of birth	30

Inverse propensity score weighting

An obstacle to updating the training data was that AHA scores are currently used to prioritize people for PSH and RRH. If being assigned housing reduces the risk of the observed outcomes, then in the future the model may erroneously correlate risk factors at the time of assessment with positive future outcomes. Simultaneously, the population which ends up being assigned to PSH or RRH may be very different from the one that does not, so simply excluding the treated population from training would create a training cohort that is fundamentally different from the population for which the model will make predictions.

We addressed this issue in two ways. First, we trained the component models only on the subset of the population that was not assigned housing. Training the models only on the untreated individuals gives the scores from the component models the intuitive interpretation "risk of outcome if not assigned to housing."

Second, we used inverse propensity score weighting (IPW) to mitigate the selection bias invoked by removing treated individuals from the training cohort. Using features from the assessment (AHA score, alt-AHA score if one exists, chronic homelessness status, belonging to any prioritized class, caseworker ID), we estimated the probability that a given assessment would be assigned to RRH or PSH. Each row was then weighted such that individuals who had a high probability of assignment were given greater weight. By upweighting individuals in the training set who were similar to those who were treated, we ensured that the model learned something about individuals with the highest levels of risk.

Further details on the assignment model and its evaluation can be found in **Appendix Table A1**.

Training methodology

We began by using LASSO to select features for each component model. A summary of the features selected for each model is shown in **Table 5**. Each component model was trained on the selected features using a random forest. The random forest models' hyperparameters were tuned using five-fold subject-wise cross-validation.

The 10-point composite scores were generated using the same procedure as the current version of AHA. In this updated version, the Any Homelessness model contributes 50% of the score and each of the other outcomes has one-sixth weight. **Figure 2** below describes how the models' scores create the AHA score.

FIGURE 2: **Mapping of Individual PRM model scores to AHA scores**

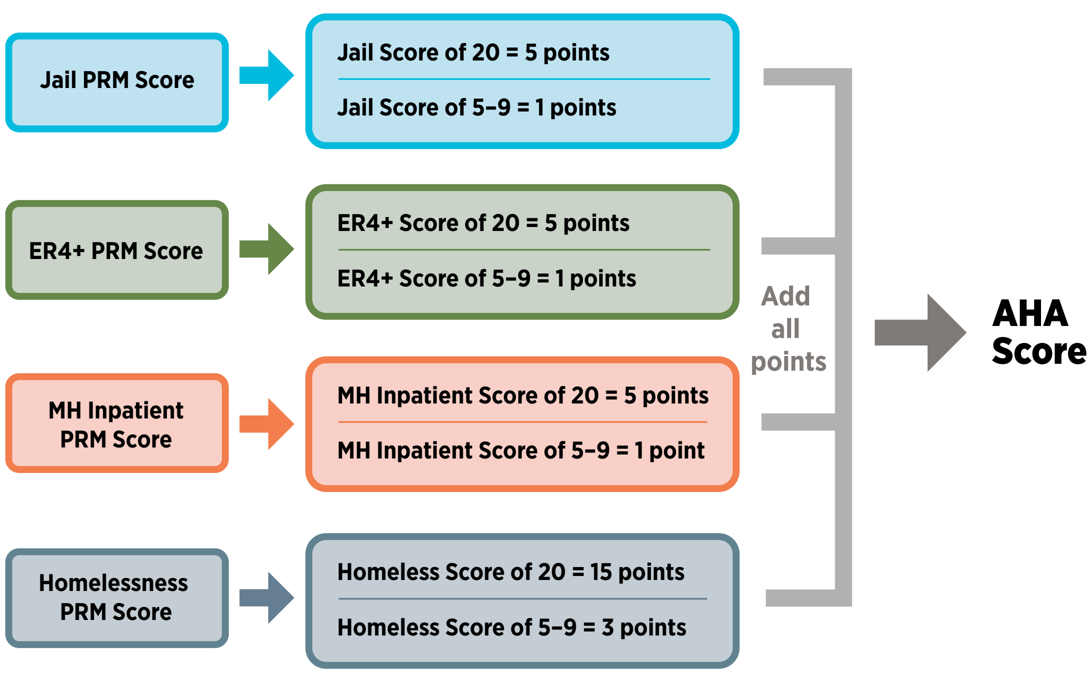


TABLE 5: **Counts of selected model features by category**

CATEGORY	MENTAL HEALTH INPATIENT	EMERGENCY ROOM 4+ VISITS	JAIL BOOKING	ANY HOMELESSNESS	TOTAL CONSIDERED
Housing and homelessness	8	5	8	22	51
Criminal/legal	19	21	67	41	246
Behavioral health	60	39	55	49	316
Physical health	9	13	10	6	54
Child welfare	7	6	2	11	72
Benefits	7	6	3	3	27
Demographics	26	25	27	29	30

PERFORMANCE EVALUATION

AHA component models

Component model performance is shown in **Table 6**. All metrics were calculated using out-of-sample (OOS) scores from a five-fold cross-validation. The AUCs for component models ranged from 0.71 to 0.83, demonstrating that all four models effectively ranked assessments by their actual level of risk. Precision and recall measures varied substantially between models in part because of differing base rates, but all models demonstrated substantial predictive power over a random draw.

TABLE 6: **AUC, precision and recall of each component model**

MODEL	AUC	PRECISION FOR 10TH RISK DECILE	RECALL FOR 10TH RISK DECILE	BASELINE PREVALENCE
Mental Health Inpatient	0.79	48%	39%	12%
Emergency Room 4+ Visits	0.75	72%	30%	24%
Jail Booking	0.83	58%	37%	16%
Any Homelessness	0.71	65%	20%	32%

Table 7 shows each component model's AUC across all four examined outcomes. As expected, each model was best at predicting its respective outcome and thus the highest AUC values are seen along the diagonal. The fact that all AUC values were above 0.5 demonstrates that these different outcomes do tend to be positively correlated with each other. Simultaneously, the high performance of each model on its specific outcome relative to other outcomes shows that each component nominates different candidates for treatment.

TABLE 7: **AUC of each model's Out-of-Sample (OOS) scores for each outcome**

MODEL/OUTCOME	MENTAL HEALTH INPATIENT	EMERGENCY ROOM 4+ VISITS	JAIL BOOKING	ANY HOMELESSNESS
Mental Health Inpatient	**0.79**	0.70	0.63	0.61
Emergency Room 4+ Visits	0.66	**0.75**	0.59	0.57
Jail Booking	0.63	0.65	**0.83**	0.63
Any Homelessness	0.55	0.55	0.59	**0.71**

AHA composite model

Table 8 shows AUCs, precisions and recalls for each outcome using the AHA composite score. The composite model showed strong performance across all outcomes of interest. People who received an AHA score of 10 were approximately twice as likely as the population average to experience any one of the four component model outcomes, thus treating just 10% of the population would allow us to intervene on over 20% of the harms observed.

TABLE 8: **AUC, precision and recall for AHA composite model**

OUTCOME	AUC	PRECISION FOR AHA 10S	RECALL FOR AHA 10S	BASELINE PREVALENCE
Mental Health Inpatient	0.71	29%	26%	12%
Emergency Room 4+ Visits	0.66	46%	21%	24%
Jail Booking	0.72	39%	22%	16%
Any Homelessness	0.68	58%	16%	32%

VALIDATION AND ROBUSTNESS

External validation

The AHA model is used to prioritize people for limited housing resources, so the best metric of model performance is precision, which tells us how many of the prioritized households would experience each of the outcomes in the absence of treatment—in this case, housing support. In addition to the four outcomes, we looked at how well the composite model identified people at high risk of fatal overdose or death.

Table 9 compares the precisions of people classified as high-risk by the updated model and the current model. We saw similar performance across the three original outcomes as well as fatal overdose and all-cause mortality. The updated model performed much better on homelessness, which is to be expected as that is the new addition to the algorithm.

TABLE 9: **Precisions of updated and current AHA, by outcome**

OUTCOME	PRECISION OF UPDATED AHA 8-10S	PRECISION OF CURRENT AHA 8-10S
Mental Health Inpatient	17%	15%
Emergency Room 4+ Visits	34%	34%
Jail Booking	28%	28%
Any Homelessness	56%	41%
Fatal Overdose	2.4%	2.5%
All-Cause Mortality	4.4%	4.5%

Fairness and equity

In August 2020, AHA replaced the VI-SPDAT as the primary tool for assessing need for housing services in Allegheny County. AHA was subject to a thorough fairness and equity review by Eticas, which found that there were few concerns regarding the model's fairness across various groups. This represented an improvement in the racial equity of housing services in the County, as research had identified racial bias in VI-SPDAT. In the process of updating AHA, we wanted to ensure that resources would continue to be allocated equitably.

This section investigates the distributional changes of moving to the updated AHA from the current version. This analysis is similar to the fairness analysis in **Tables 17 and 18** of the September 2020 Methodology Report, which showed how moving from VI-SPDAT to AHA would change the race and gender distribution of singles and families assigned to RRH/PSH.

Those most likely to be assigned to RRH and PSH are families with a score of seven or above and singles with a score of 10. **Tables 10 and 11** show the differences in the racial and gender distribution of single AHA 10s and family AHA 7–10s between the current and updated model to show how updating the model would impact the allocation of housing.

Allocations across racial groups appear similar in the updated model, but there is a shift in allocations between genders. In the new model, more men and fewer women will be assigned housing for both singles and families. This shift is likely due to the large difference in homelessness risk between men (38% one-year homelessness risk) and women (26%). Because the updated model includes the homelessness component, it prioritizes more men for housing.

In summary, the results from this section suggest that updating the AHA model does not substantially shift the allocation of resources.

TABLE 10: **Change in distribution of AHA 10s, singles**

	UPDATED MODEL	CURRENT MODEL
Black	53%	52%
White	46%	45%
Female	23%	30%
Male	77%	68%

TABLE 11: **Change in distribution of AHA 7-10s, families**

	UPDATED MODEL	CURRENT MODEL
Black	56%	58%
White	43%	40%
Female	71%	83%
Male	29%	17%

Live performance

The model performances shown so far were based on OOS scores, but it is also important to understand how models perform in a production setting. We scored AHA assessments from March 2025 to May 2025 using the updated AHA model and looked at the precision with which a score of 10 predicted an outcome within two weeks. We focused on the Jail Booking and Any Homelessness outcomes because there is a multi-month lag in obtaining the claims data necessary to look at the short-term precision for the other two outcomes.

Table 12 shows the results from this comparison. The two-week base rates were similar for the training and validation cohorts. The updated AHA model showed similar performance in production to the performance observed in training and consistently outperformed the existing model on these two metrics.

TABLE 12: **Two-week precisions for Jail Booking and Any Homelessness outcomes**

	JAIL BOOKING		ANY HOMELESSNESS	
METRIC	PRECISION AHA 10S	BASE RATE	PRECISION AHA 10S	BASE RATE
Training cohort, updated AHA	3.8%	1.1%	13.7%	9.3%
Validation cohort, updated AHA	4.0%	1.4%	19.0%	8.3%
Validation cohort, current AHA	2.5%	1.4%	15.2%	8.3%

Impact on Related Model: MH-AHA

The Mental Health – Allegheny Housing Assessment (MH-AHA) is used to help determine criteria for mental health residential (MH Res) programs. MH-AHA was constructed using two of the components of the AHA model, specifically the Mental Health Inpatient and Emergency Room 4+ Visits. Because of this interdependence, any update to AHA would simultaneously update MH-AHA. We ask readers to refer to the November 2023 Methodology Report for additional background and context regarding MH-AHA and MH Res.

To verify that the updated MH-AHA performed just as well as the version in production, we constructed a validation dataset containing every historical MH Res referral up through February 2024. This dataset included individuals' current MH-AHA score along with their updated MH-AHA score. Of the 4,637 referrals in these data, about 82% were made before the implementation of MH-AHA and required the computation of retro scores.

Table 13 shows the AUCs of the updated MH-AHA compared with the current version. We observed improvement in the AUCs for both outcomes in the validation data. This result reflects the overall improvement in the Mental Health Inpatient and Emergency Room 4+ Visits models shown in **Table 2**.

TABLE 13: **AUC of updated MH-AHA vs. current MH-AHA, by outcome**

OUTCOME	AUC OF UPDATED MH-AHA	AUC OF CURRENT MH-AHA
Mental Health Inpatient	0.73	0.64
Emergency Room 4+ Visits	0.78	0.75

Table 14 compares the precision and recall between the updated MH-AHA and current MH-AHA for individuals with a score of 9 or 10. The updated MH-AHA had higher precisions and recalls than the current model across all outcomes.

TABLE 14: **Metrics for high-risk individuals from updated and current MH-AHA**

	MH-AHA 9-10S	UPDATED MH-AHA	CURRENT MH-AHA
	% of Population	17%	16%
Mental Health Inpatient	**Precision**	48%	40%
	Recall	34%	28%
Emergency Room 4+ Visits	**Precision**	52%	45%
	Recall	48%	41%

AUTHOR/ANALYTICS:
Steven Lann, PhD , Erika Montana and Pim Welle, PhD, Allegheny County Department of Human Services

REVIEWERS:
Rhema Vaithianathan, PhD (Auckland University of Technology) and Bryan Wilder, PhD (Carnegie Mellon University)

FOUNDATIONAL WORK:
The Center for Social Data Analytics at the University of Auckland created the AHA model, and Rhema Vaithianathan led its implementation. Since August 2020, the model has been used to prioritize individuals at high risk for Allegheny County's limited rapid rehousing and permanent supportive housing units.

This report describes the most recent model update. As of this update, DHS staff and systems will directly maintain, update and house the model.

SUPPORTING STAKEHOLDERS
The following people were key contributors in developing, reviewing and implementing this model at ACDHS: Nicolas Marlton, Felipe Diaz, Diana Martinez, Ethan Goode, Abhishek Jallawaram, Honey Rosenbloom, Zach Kurtz.

Additionally Andrea Krivosh, Colleen Cain, Rachel Rue, Dinesh Nair and John Villela provided subject matter expertise.

APPENDIX

Assignment Model

The Assignment Model predicts the probability that a given RRH/PSH assessment will result in assigned housing. The training data consists of a subset of the universe of assessments; we used several rules to select data for the model to ensure data quality. First, we filtered out "soft assessments" where applicants were not asked the full list of assessment questions. Next, we only kept assessments that were at least 60 days from a prior assessment. This prevented the inclusion of multiple assessments for the same person in a short period of time, which is particularly an issue in the pre-AHA era where we routinely observed several assessments in a span of a few days or weeks for a single individual. These rules gave us a training cohort of 27,709 assessments from 17,676 individuals, with about 13% of these assessments leading to assignment.

We used the year of assessment, SPDAT or AHA score, alt-AHA score if one existed, chronic homelessness status, belonging to any prioritized class and caseworker ID as model features. The model was trained using a random forest that was tuned using five-fold cross-validation. The performance metrics for the model are reported in **Table A1**. The AUC of 0.88 indicates that the model effectively ranks assessments by their likelihood of assignment.

TABLE A1: **AUC, precision and recall for RRH/PSH Assignment Model**

OUTCOME	AUC	PRECISION FOR TOP 10%	RECALL FOR TOP 10%	BASE RATE
Assignment to RRH/PSH	0.88	56%	42%	13.3%