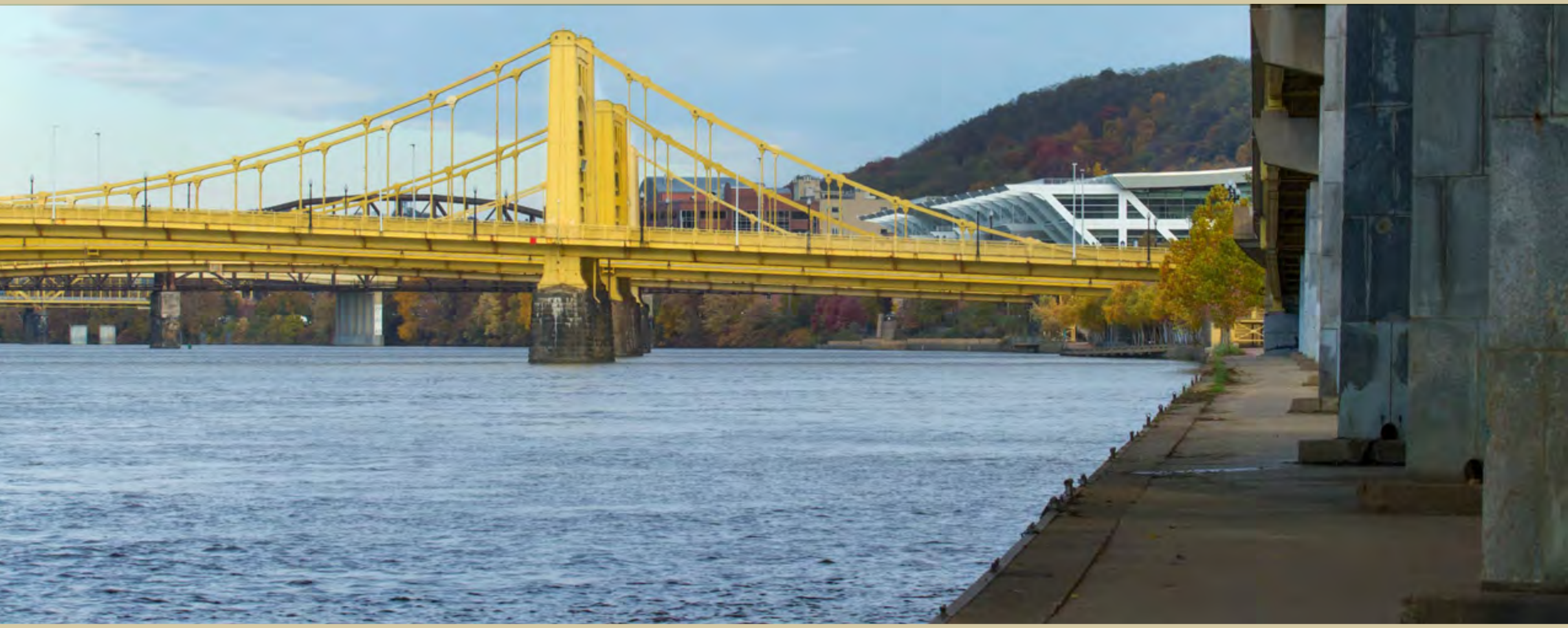
Introducing the Allegheny County Street Stabilization Team



Prepared by the Allegheny County Department of Human Services

AUGUST 2026

The Allegheny County
Department of Human Services
One Smithfield Street
Pittsburgh, Pennsylvania 15222

analytics.alleghenycounty.us

TABLE OF CONTENTS

PROJECT OVERVIEW

Introducing The Street Stabilization Team

The Street Stabilization Team (SST), a new intervention launched by the Allegheny County Department of Human Services (ACDHS) in July 2026, was designed to address the difficulty healthcare systems and publicly funded programs face when supporting people who demonstrate profound disability and/or high vulnerability and are chronically disengaged from clinic-based care due to the severity of psychiatric symptoms including but not limited to serious mental illness (SMI) and substance use disorder (SUD). These Allegheny County residents have frequent interactions with first responders (including law enforcement) and high utilization of crisis services (e.g., homelessness street outreach, mental health crisis response, emergency department visits, involuntary psychiatric hospitalizations). In many cases, members of this group are well known to service providers in the County, but existing services have not been successful in meeting their complex needs.

In response, ACDHS launched a comprehensive stakeholder input process in 2024 involving practitioners and leaders in homelessness services, street medicine, inpatient psychiatry and behavioral health crisis response, along with healthcare counterparts working with these residents when they appear at court and while in the Allegheny County Jail (ACJ). Over the next year and a half, this group—which came to be known as the Acute Working Group—identified key challenges that interfere with their ability to serve the County's most vulnerable unhoused individuals. Ultimately, they recommended the creation of a multidisciplinary, street-based outreach team led by a psychiatrist and designed to integrate care across the region's two major hospital systems and County-funded programs. SST is the result of that process.

SST aims to serve 20–30 people who have not meaningfully engaged with traditional treatment supports due to a combination of factors, including the severity of their behavioral and other health needs and gaps in services provided by the health and human services systems. In order to be eligible for SST a person must:

- Be an Allegheny County resident over the age of 18
- Be currently unhoused, most likely street homeless
- Demonstrate severe psychiatric symptoms, including but not limited to those due to traditional SMI and SUD diagnoses
- Be unwilling or unable to engage in traditional services due to severity of symptoms
- Be identified as high vulnerability based on a street assessment and/or risk modeling

The 2026 launch of SST is a pilot; after we evaluate the efficacy of the intervention in its first year, we will explore options for expanding to a larger cohort-based on need.

The Mission of SST

SST brings together staff from ACDHS, University of Pittsburgh Medical Center (UPMC) and Allegheny Health Network (AHN). The work is guided by a shared mission:

> *To provide clinical treatment and care coordination for unhoused individuals with multi-system needs, frequent crisis encounters, and high risk of severe adverse outcomes who are currently not served by existing systems. Care will be provided in a manner that is timely, comprehensive, collaborative and equitable.*

Designing SST

In 2024, ACDHS began convening stakeholders from regional homelessness services programs, healthcare networks and crisis response teams to participate in conversations about a clear gap in the County's services for unhoused individuals with severe and co-occurring behavioral and physical health issues. Many of the people coming to the attention of these stakeholders were also regularly navigating interactions with law enforcement, incarceration and involuntary hospitalization. Despite their frequent crisis encounters and the best efforts of first responders and mental health mobile crisis teams, these individuals were not meaningfully accessing longer term supports. The Acute Working Group laid out case conceptualizations for those identified as most impacted and defined a set of discreet barriers preventing their more successful engagement with community resources. These barriers became the foundation for the group's subsequent collaborative design of the SST intervention.

Barriers

The Acute Working Group identified the following barriers:

- Insufficient identification of and shared focus on unhoused residents living with severe and unmet behavioral health needs
- Insufficient real-time communication between ACJ, street outreach teams, crisis response teams and emergent care providers during clients' crisis events (i.e. sudden or escalating situations that require immediate intervention beyond routine treatment or supportive services)
- Insufficient information sharing between ACJ, street outreach teams, crisis response teams and inpatient providers
- Lack of frontline workers assigned or scoped to persistently engage and support unhoused residents with severe and unmet behavioral health needs
- Difficulty verifying eligibility for existing treatment teams can exclude patients with complex needs that challenge existing treatment paradigms
- Payment and regulatory structures for existing treatment teams that discourage persistent engagement practices

In response, we designed SST to be:

Cross-system and Team-Based

The population that SST serves touches multiple providers and systems across the County. One goal of SST is to break the silos that separate different elements of client care. A cross-functional team, led by a psychiatrist from Western Psychiatric Hospital (WPH) and staffed by caseworkers from ACDHS, a street nurse from AHN and an internal medicine physician from AHN, will provide coordinated services. This team will be able to advocate for clients across providers and navigate systems on their behalf.

Low-Barrier and Recovery Oriented

SST is grounded in a recovery-oriented philosophy that prioritizes the goals, preferences and lived experience of each individual. Rather than requiring clients to fit into existing service structures, SST adapts engagement, planning and interventions to align with each person's circumstances and readiness for change. The approach emphasizes dignity, autonomy and collaborative decision-making. Harm reduction strategies are central to the model, allowing the team to support incremental progress and improved health and safety without imposing rigid participation requirements. Low-barrier engagement ensures that individuals with significant mistrust of systems, inconsistent service use and/or ambivalence toward treatment can still access meaningful support.

Persistent and Community-Based

The SST model was intentionally designed to be persistent, flexible and rooted in community settings rather than traditional office-based environments. The team practices assertive engagement and begins engagement before enrollment. Participation is not dependent on a formal opt-in process; instead, the team proactively connects with clients in the locations where they live and spend time. By maintaining a steady presence and building trust over time, SST can respond rapidly to emerging needs and coordinate across systems to ensure that clients are not left without support during critical moments.

Where SST Fits in with Other Services

SST was designed to address a gap in services in Allegheny County by serving a specific group for whom other interventions have not worked. Although alternative programs that complement SST exist, the Acute Working Group determined that these services either serve a narrower population with less severe unmet behavioral health needs or are not structured to meet the complex needs of SST-eligible people.

Still, other similarly comprehensive services use models comparable to SST. To avoid duplicating support, individuals enrolled in certain services, as described here, will generally not be eligible for SST.

Community Treatment Team (CTT)

CTT is a community-based intervention that delivers comprehensive mental health services to individuals with SMI, particularly schizophrenia spectrum disorders. Through its multidisciplinary teams offering around-the-clock support and integrated care, CTT plays a crucial role in reducing avoidable hospitalization and enhancing client stability.

CTT and SST serve different populations. CTT is an intervention designed for individuals with SMI; SST was designed for people who are both unhoused and have severe unmet behavioral health needs. SST is a response to the reality that programs like CTT are not sufficient to meet the needs of some County residents. In general, if someone can actively participate in CTT, they are unlikely to need the level of service offered by SST. Therefore, anyone actively participating in CTT is not eligible for SST.

Integrated Dual Disorder Treatment (IDDT)

IDDT is a community-based, evidence-driven program targeting individuals with co-occurring mental health diagnoses and SUD. Its strengths lie in integrated, stage-specific interdisciplinary care for adults experiencing these co-occurring conditions who can benefit from treatment that is integrated rather than siloed between separate care systems. Within the County's adult behavioral health continuum, IDDT sits alongside CTT, focusing specifically on co-occurring disorders within a community-based treatment framework.

IDDT requires a SMI diagnosis for eligibility and does not have a specific focus on the unhoused population. Because it operates on a fee-for-service model—billing in 15-minute increments—and treats unsuccessful attempts at engagement as non-billable, it is not designed for ongoing outreach and engagement efforts. By design, SST encourages attempted outreach and does not require a specific diagnosis as an eligibility criterion. That said, there are people who qualify for both SST and IDDT. Someone actively participating in IDDT is not eligible for SST.

Permanent Supportive Housing (PSH)

PSH in Allegheny County is a long-term, housing first–oriented program designed for individuals with significant behavioral health needs who require stable housing paired with ongoing support. It provides permanent subsidized rental units linked to voluntary services that promote recovery, wellness and community integration. PSH serves people who have experienced chronic homelessness and/or repeated institutional stays and who, without sustained assistance, face substantial barriers to maintaining housing. Within the County's broader housing and behavioral health continuum, PSH functions as a high-stability option for those needing intensive and flexible wraparound supports to remain safely and successfully housed in the community.

SST is designed to support people who are both actively unhoused and managing co-occurring conditions. Stable enrollment in PSH would mean an individual was not eligible for SST.

Rapid Re-Housing (RRH)

RRH is a short- to medium-term intervention aimed at helping individuals and families exit homelessness quickly and attain stability through time-limited financial assistance and targeted case management. Grounded in Housing First principles, RRH prioritizes reducing the length of homelessness by rapidly connecting people to safe, market-rate housing while offering support such as rental assistance, move-in funding and housing navigation. It serves households with moderate service needs who can achieve long-term stability once immediate barriers are removed. Within the County's Continuum of Care, RRH is intended to act as a bridge to longer-term options like PSH. RRH enrollment also impacts SST eligibility.

Other Cases That May Preclude SST Eligibility

SST exists to serve individuals who are both unhoused and not maintaining consistent engagement with other services while dealing with severe and unmet behavioral health needs. Eligibility, ability and willingness to participate in other intensive services may indicate that a person's situation does not match the case conceptualization of the population SST is designed to serve. For this reason, people who are maintaining consistent participation in services like Medication-Assisted Treatment, intensive case management and/or outpatient therapy are also likely to be ineligible for SST.

REFERRALS, EVALUATIONS AND SERVICE DELIVERY

SST Services

Participants in the SST program receive a comprehensive blend of behavioral health, medical and social supports delivered by a coordinated, multidisciplinary team. The program provides mobile diagnostic assessments, medication management and access to both psychiatric and basic medical care (e.g., wound treatment, lab coordination, support for medication adherence).

Participants benefit from harm reduction–oriented outreach, persistent engagement and help navigating housing resources, public benefits and legal or court processes. The team also offers health education, facilitates communication with hospitals and other service systems and ensures coordinated risk assessment and crisis response. Together, these supports create a flexible, community-based intervention that meets individuals where they are and helps reduce barriers to stability, treatment and recovery.

SST Roles and Engagement Model

SST employs a multidisciplinary team with a shared caseload, by design, to create a more resilient and responsive service model. This shared approach shields clients from developing unsustainable relationships with a single staff member, enabling high-contact and flex -ible engagement regardless of individual staff availability. It also ensures continuous coverage during absences and promotes collective clinical accountability. Importantly, by distributing responsibility, the design also maintains consistent team awareness of each client's needs.

Still, SST is a specialized intervention with clear distinctions between roles and functions. The roles and responsibilities of the six team members are detailed in Table 1.

TABLE 1: Street Stabilization Team Member Roles and Responsibilities

ROLE	PRIMARY CLINICAL FUNCTION	SHARED RESPONSIBILITIES
Psychiatrist	Clinical assessment, treatment planning	Clinical decisions and treatment lead for all clients
Street Nurse	Administer medication and medical care	Health monitoring for all clients
Consulting Physician	Address complex and chronic physical health conditions, consult with other professionals and team members	Physical health consultation for all clients
Social Worker A	Housing and benefits lead	Engagement for all clients
Social Worker B	Court and systems lead	Engagement for all clients
Program Director	Oversight and escalation	Risk review for all clients

SST Referral Pathways

Because SST eligible clients are managing severe and unmet behavioral health needs and are routinely disengaged from human services programs, there is no single referral process that can be expected to identify all eligible individuals. For this reason, SST accepts referrals made through one or both of two channels. The first channel is provider referrals made by a network of medical practitioners, homelessness services providers, crisis response teams and others who interact with people who might benefit from SST. The second channel, *data-driven referrals*, is a data-driven model that utilizes individual-level administrative data to assess the acuity of individual cases. Using both processes allows us to build a referral list that reflects provider insights missing from administrative data and service use patterns that providers may not see.

All referrals are subject to a review by a caseworker and a clinician. The full team makes consensus-based prioritization and enrollment decisions after referrals from either channel pass these rounds of human review. To be eligible, an individual must be an Allegheny County resident who is 18 or older, currently unhoused and demonstrate profound disability and/or high vulnerability and chronic disengagement from clinic-based care due to the severity of psychiatric symptoms including but not limited to SMI and SUD.

Provider Referrals

Provider referrals are a critical entry point into the SST program. Many of the people served will be familiar to frontline workers who encounter people experiencing repeated crises, severe psychiatric symptoms, unsafe living conditions and/or significant barriers to traditional services on a daily basis. This referral pathway ensures that these individuals can be identified by the frontline teams who know them best. The following providers are recognized referrers for SST:

- City of Pittsburgh Office of Community Health and Safety providers
- resolve Crisis Center
- Street Medicine teams
- Street Outreach teams
- Alternative Response Team (A-Team)
- Complex case meeting for incarcerated individuals

Data-Driven Referrals

Data-driven referrals use administrative data by combining estimates of the risk of incarceration, involuntary hospitalization, shelter use and fatal overdose for a given individual. It combines these estimates into a single assessment, then filters out those who are already engaged in CTT, IDDT, PSH, RRH and/or other intensive services.

The data-driven model creates a referral list of 100 eligible people each month. There is significant overlap between the data-driven referrals each month. So, over a one-year period, we expect the data-driven process to refer a total of 250-300 distinct people. There is also significant overlap between data-driven referrals and provider referrals. In general, we expect that 60%-70% of provider referrals will appear in the model referral list.

Evaluation and Enrollment in SST

After an individual is identified through either referral pathway, the SST enrollment process begins with a case conference in which the team reviews available information to determine whether the person's needs align with the program's eligibility criteria. This chart review process uses data from a variety of Electronic Health Record (EHR) systems and Homeless Management Information Systems (HMIS). The reviews are augmented with summaries of ACDHS administrative data. These summaries are produced using a Large Language Model (LLM) as detailed later in this report (see Data-Driven Decision Support).

Next, a psychiatric screening is administered to assess clinical acuity and guide initial care planning. The team makes prioritization and enrollment decisions through team consensus to ensure a shared understanding of risk, needs and appropriate next steps. Once the team determines that an individual is eligible, they notify them verbally and obtain consent as soon as possible. In the case of provider referrals, the SST team will notify the referring provider whether an individual has been accepted for service or not.

If the person is unable or unwilling to provide consent at that stage, the team will begin outreach and engagement and maintain persistent contact to build rapport in the hope that the person will decide to formally enroll. If someone does not respond to outreach over a period of multiple months, the team may decide to pause active service and move the individual onto a waitlist.

Engagement and Outreach

SST uses an assertive engagement model that prioritizes meeting clients where they are—both physically and psychologically—and building trust over time. Engagement begins before any formal enrollment or paperwork and emphasizes rapport, tangible support such as food or wound care and rapid field-based clinical assessment. The team can initiate certain psychiatric and medical interventions directly in the field, when warranted and safe, including medication and long-acting injectables. All engagement is grounded in trauma-informed care principles that assume high trauma prevalence in unsheltered populations and seek to avoid re-traumatization through clear communication, respect for personal space, shared decision-making and use of non-judgmental language. Importantly, substance use is never a barrier to care; efforts focus first on reducing mortality through tools like naloxone distribution, safer-use education and low-threshold buprenorphine initiation.

Engagement is persistent and designed to withstand missed contacts, refusals or periods of instability. Outreach frequency is matched to clinical risk, with increased contact occurring during transitions (e.g., from hospital, ACJ exit). The team uses an engagement ladder, moving flexibly from simple presence to deeper clinical conversations and full care-coordination as trust develops. Documentation of all encounters and attempted contacts is to be done within 24 hours. Consistent with a non-termination philosophy, no client will be discharged for substance use, missed appointments, refusals or behavioral dysregulation except in rare cases of unmanageable violence.

Discharge occurs only when a client relocates permanently or voluntarily transfers care, or when risk exceeds the team's capacity; a structured transition plan is required in each case.

DATA-DRIVEN DECISION SUPPORT

Data-driven decision support is the application of a predictive model as an input to a human decision-making process. A routine example is the use of a computer vision model in radiology. In that case, the model makes a diagnosis and so does the physician. The physician then makes a recommendation on the basis of both the model and human diagnoses.

SST is designed to use a data-driven model for decision support. The data-driven model used in the referral process provides a list of names collated with a list of names from provider referrals. The decision support here is the addition of people to the list who would be missed if not identified by a provider.

Caseworkers and clinicians use a variety of systems and methods to inform their decision making (e.g., EHR systems, HMIS, conversations with peers). For SST, one additional source is a text summary—written by the LLM—that outlines an individual's interactions with County programs. The process extracts and aggregates the data using simple, deterministic code. The sole purpose of the LLM is to translate the content from tabular data into natural language summaries. Reviewers have access to the raw data in addition to the summaries and can dive deeper if they choose.

See Language Model-Supported Case Conceptualization for more information about creation of the summaries by LLM.

Why Use a Data-Driven Model to Make Referrals?

SST was designed to serve people who have not maintained meaningful engagement in County services, meaning that there are eligible individuals who are not familiar to the provider network. Thus, only asking providers for referrals risks privileging people with more recent system involvement and more established relationships with providers. The data-driven model uses risk of four outcomes (homelessness, incarceration, involuntary hospitalization and fatal overdose) as the basis for referrals. We selected these outcomes to approximate the co-occurring risks facing individuals who are eligible for SST (further detail on each outcome is available in **Table A2 of Appendix B**). Using only the data-driven model risks privileging people with more prominent administrative data. By combining the two pathways, we include both those known to the provider network and those at high likelihood of negative outcomes who are consistently disengaged from clinic-based care.

How Using LLMs Supports Case Conceptualization

During the caseworker and clinical reviews, SST members conduct a detailed assessment of each person referred. Because SST is a collaboration between UPMC, AHN and ACDHS—each with its own data systems—each assessment can be extremely time consuming. SST uses the LLM data summarization tool to aggregate ACDHS data into easy-to-read case conceptualizations. The process has two parts: data are extracted from relevant systems and then the LLM puts the data into a human-readable format.

The LLM is narrowly used—it is restricted from making inferences or recommendations beyond the procedural summarization of data that would otherwise be impractical for a caseworker or clinician to collect and review.

METHODOLOGY

Predictive Model Referral Process

How SST Uses the Data-Driven Model to Make Referrals

As mentioned previously, the goal of the model is to refer eligible people for SST. The data-driven referral list is created using a composite of predictions from four data-driven models (one for each of four potential negative outcomes) and filtered based on the set of eligibility criteria previously described. The data-driven referral list includes 100 currently or recently unhoused individuals (known within ACDHS's administrative data) with the highest composite rating who meet the eligibility criteria.

SST staff and clients are never shown the raw composite rating or component-model outputs. Any individual excluded from the data-driven referral list for any reason is still eligible for a provider referral.

The composite is a deterministic function of the outputs of four data-driven models that predict binary outcomes (0,1):

- Involuntary hospitalization (302): Equal to 1 if a person has a 302 petition that is upheld in the year after the referral date
- Jail booking at ACJ: Equal to 1 if a person has a booking at ACJ in the year after the referral date
- Shelter system interaction: Equal to 1 if a person checks into a County shelter, continues to stay in a shelter for at least 60 days or interacts with a street outreach team in the year after the referral date
- Fatal overdose: Equal to 1 if the Medical Examiner determines that a person died from an overdose in the year after the referral date

The composite is produced by summing the standardized outputs of these four models after adjusting weights by outcome. It is used as a proxy for co-occurring severe psychiatric symptoms, chemical dependency and acute risk of morbidity and mortality. We prefer the use of a composite over predicting the union of the four outcomes because the composite better approximates the occurrence of multiple outcomes while the union would effectively approximate the probability of any one of the outcomes.

Outcome definitions can be found in **Appendix B, Table A2**. For information about composite methodology, see *Composite Model Methodology* in the **Appendix**.

Model Input Data

Our data come from Allegheny County's data warehouse and include information on involvement in the behavioral health, child welfare, criminal-legal, homelessness and other service systems. A detailed account of the features used in the model can be found in **Appendix B, Table A4**.

While the data cover a wide variety of domains, there are limitations. The data warehouse largely contains information on engagement with human services programs for Allegheny County residents. Individuals who recently moved to the County or who have not utilized services may not be included in the data (or their information may be incomplete).

Another limitation is that we only have access to data about physical and behavioral health services paid for by Allegheny County, Medicaid or UPMC insurance; services paid for privately or through other private insurance may not be available in the data.

We intentionally exclude information older than five years to uphold an individual's right to be forgotten; we have not observed any significant decrease in model performance by limiting the time horizon in this way.

We trained the models on data from 531,269 individuals with ACDHS service involvement in the five years preceding 6/10/2024. Each row in the training data represents a person's interaction with ACDHS services over that five-year period. Outcomes were flagged as binary (0,1) columns for the year starting 6/10/2024. A detailed breakdown of the demographics of the training data can be found in **Appendix B, Table A5**.

Table 2 shows the base rates for the predicted outcomes in the training data.

TABLE 2: **Data-Driven Model Predicted Outcomes**

OUTCOME	FREQUENCY	BASE RATE IN THE POPULATION
Involuntary hospitalization (302)	2,509	0.47%
Jail booking at ACJ	5,870	1.10%
Shelter system interaction	2,515	0.47%
Fatal overdose	338	0.06%

Component Model Performance

The models return risk assessments that are decimals between 0 and 1, where a higher assessment means a higher expectation that the outcome will occur in the following 12 months. Because these outputs are used to order a list of possible referrals, we need to be confident that individuals with higher model assessments have higher incidence of the predicted outcomes.

Table 3 shows the assessment of the performance of each outcome model based on five-fold out-of-sample predictions as of 6/10/2024. These test results suggest that the outcomes occur at significantly higher rates for the people with the 1,000 and ~5,000 highest assessments from each of the component models. Precision in this context means the rate at which an outcome happens in a given group. Comparing precision to the base rate of an outcome tells us how much more frequently we observe that outcome in a subgroup than we do for the full population.

For each of the four component models, we report precisions for the top 1,000 assessments that are at least an order of magnitude higher than the base rate in the population. In all cases, those precisions are at least 50 times higher than the base rates.

TABLE 3: **Data-Driven Model Performance by Outcome**

METRIC	302	JAIL	SHELTER	OVERDOSE
Precision (Top 0.2%, n=1000)	34.7%	62.3%	70.1%	3.3%
Compared to base rate	73.5X	56.4X	148.1X	51.9X
Precision (Top 1%, n=5,312)	16.2%	40.4%	25.9%	2.1%
Compared to base rate	34.3X	36.6X	54.7X	33.4X

Quantitative Assessments of Composite Rating Quality

The referral list includes the subset of people who have the highest composite ratings based on the model assessments after filtering for eligibility. Each month, 100 people are referred by the model. More than two-thirds of the people referred in a given month will still be on the referral list the next month.

Considering outcomes for the top 50, 100 and 200 people on the list (after filtering for eligibility) helps describe the severity of the situation of the people who are referred. Importantly, each row in **Table 4** describes outcome rates for a single group of people. By contrast, the precision metrics in **Table 3** used a different cohort for each outcome and n-size.

In the year after their referral date, we expect more than 90% of people on the referral list to experience one of the four predicted outcomes. In most cases, they will experience more than one.

TABLE 4: **Composite Model Performance by Outcome**

METRIC	ACJ	OVERDOSE	302	SHELTER	ONE OR MORE
Precision (Top 50)	66.0%	4.0%	34.0%	76.0%	94.0%
Precision (Top 100)	63.0%	3.0%	23.0%	68.0%	89.0%
Precision (Top 200)	54.5%	1.5%	16.5%	62.0%	80.5%
Base Rate Among Unhoused, Eligible Residents	16.4%	0.5%	4.1%	40.8%	48.4%

A practical way to benchmark the performance of a data-driven model is to compare it to a reasonable heuristic (or "rule of thumb"). To make this comparison, we ran our referral process once per month on historical data for each month of 2024. For each month of 2024, we also selected a cohort of people using a heuristic. The heuristic required that a person was eligible for SST and that they had each of the following four events in the year preceding the referral date: interactions with the shelter system, at least one booking in ACJ, a County-paid or Medicaid claim for SUD treatment and a similar claim for behavioral health services. The model returned 1,200 referrals (representing 284 unique individuals) and the heuristic returned 1,885 referrals (representing 339 unique individuals). More details on the composition of the heuristic referral cohort are available in **Appendix B, Table A15.**

Table 5 compares data for distinct referrals, not distinct individuals. Comparing performance by distinct referral allows us to observe the rate at which outcomes occur for the average referral irrespective of the number of times an individual was referred. It is worth noting that for the data-driven referral process, the outcome rate statistics would be identical if we took the simple average of the outcome rates for each of the 12 timeframes.

TABLE 5: **Composite Model Performance Compared to an Alternative Process**

-	OUTCOMES					-
REFERRAL METHOD	ACJ	OVERDOSE	302	SHELTER	ONE OR MORE	N
Composite (n=1,200)	65.2%	2.6%	23.0%	70.4%	89.8%	1,200
Heuristic (n=1,185)	46.4%	0.8%	16.3%	45.9%	70.0%	1,885
Improvement Over Heuristic (%)	40.4%	206.1%	41.2%	53.4%	28.3%	-

Readers will notice that the composite model outperformed the reasonable heuristic on all outcomes. We interpret this to indicate higher acuity among the cohort of data-driven referrals.

Qualitative Assessments of Composite Rating Quality

SST also employed a qualitative assessment method for potential lists of model referrals. Frontline workers (street nurses and crisis response teams) reviewed the names of the 30 people with the highest composite ratings as of October 2025. We used the top 30 because that was a practical number of cases to discuss in the time allotted. We asked who the outreach workers knew and who they assessed to be a fit for the program.

Of the 30 individuals, 19 were known to the outreach workers. Of the known individuals, 10 were assessed to be a fit for the program based on outreach worker experience and administrative data. The group determined that the other nine were not a fit based on engagement with other services that were not on the data-driven exclusion list. This check is not particularly well-powered statistically but gives an indication that the list includes two important groups: people who match the case conceptualization and people who would not have been referred by the providers we surveyed.

Intuitively, we believe that too much or too little agreement would be concerning. Too much agreement would signal that the models were not useful for referring individuals who may not have been included in provider referrals. Too little agreement could suggest that the models were not identifying people deemed to be high acuity by frontline workers.

Model Fairness

In addition to performance, we used standard measures to assess the fairness of our targeting allocation. The key metric we used was the ratio of recall among demographic subgroups, reported in **Table 6** as Sensitivity Ratio. Recall is the fraction of all cases among the cohort referred by the model in which one or more outcomes occur. By computing the Sensitivity Ratio, we judged whether the model assigned output fairly based on subgroup demographic. We do not expect the process to have a Sensitivity Ratio of exactly 1.0 and we consider models with Sensitivity Ratios between 0.7 and 1.3 as performing well according to this fairness diagnostic.

Table 6 compares non-White and White individuals in our data, while **Table 7** compares female and male individuals. Together, we found that the models gave similar allocations across race and gender lines and performed similarly to a comparable heuristic. With regards to race, the composite showed slightly lower recall for non-White individuals compared to a reference group of White individuals. The simple heuristic also showed lower recall among non-White individuals, to a similar degree.

TABLE 6: **Composite Model Fairness for Race, Compared to an Alternative Process**

METRIC	REFERRAL METHOD	PRECISION (NON-WHITE)	PRECISION (WHITE)	RECALL (NON-WHITE)	RECALL (WHITE)	SENSITIVITY RATIO
One or More Outcomes	Composite (n=1,200)	91.5%	88.3%	8.1%	9.4%	0.86
	Heuristic (n=1,185)	70.1%	69.1%	6.5%	7.0%	0.93

When comparing gender, we used binary male and female labels because gender data are tracked as a binary in the Allegheny County data warehouse. The composite showed slightly lower recall among females than among males. The heuristic also had lower recall for females as compared to males.

TABLE 7: **Composite Model Fairness for Gender, Compared to an Alternative Process**

METRIC	REFERRAL METHOD	PRECISION (FEMALE)	PRECISION (MALE)	RECALL (FEMALE)	RECALL (MALE)	SENSITIVITY RATIO
One or More Outcomes	Composite (n=1,200)	93.8%	88.1%	8.5%	8.8%	0.97
	Heuristic (n=1,185)	72.1%	69.2%	6.2%	6.9%	0.89

One important note for **Tables 6 and 7** is that the lower recall among the group selected with the heuristic is partially attributable to the smaller n-size (1,185) for the heuristic as compared to the composite (1,200). However, only a small part of the difference is attributable to differences in selection. For example, even if we adjusted the n-size for the heuristic to 1,200 by manually including 15 true positives, the composite still outperformed the heuristic on precision and recall by at least 15 percentage points in all cases.

Language Model-Supported Case Conceptualization

The Summaries

The summaries act as a starting point for case conceptualization. They provide plain text descriptions of a client's interactions with County services and institutions. Figure 1 shows an example report that includes synthetic data on housing services and evictions as well as behavioral health services and diagnoses. Reports used for SST have additional domains as documented below.

FIGURE 1: **Language Model Summary Example**

Client Report

This summary was made by an AI language model on May 1 2026, using data from the Allegheny County data warehouse. The data warehouse can be incomplete or out of date. The AI model may omit important details.

Client Information

ID: 0000000
FNAME: Fname
LNAME: Lname
DOB: YYYY-MM-DD

Admin Data Summary

Housing Services & Evictions
The client's first housing-related event was an emergency shelter stay in January 2019. In the past year, the client received transitional housing placement, rapid rehousing assistance, and permanent supportive housing placement, totaling three housing interventions. In the last three months, the client remained in a permanent supportive housing placement, specifically since March 1, 2026. The client is currently residing in permanent supportive housing.

Behavioral Health Services & Diagnoses
The client's first behavioral health event was receiving SUD non-opioid or OUD treatment through HealthChoices in 2003.

In the past year, the client had 4 interactions with behavioral health services, including:

- 3 mental health episodes in an emergency room
- 1 mental health episode with an inpatient visit
- 1 record of receiving SUD non-opioid or OUD treatment through HealthChoices

In the last three months, the client had 2 behavioral health events: one emergency room mental health episode and one instance of receiving SUD treatment through HealthChoices, which has been ongoing since June 1, 2025.

There are no current ongoing behavioral health events beyond those noted in the past 3 months.

Recent Events (Past 90 Days)

Description	Most Recent Event	Event Count
Mental health episode in an emergency room	2026-04-01	1
Permanent Supportive Housing placement	2026-03-01	1

How We Use Summaries

Members of the SST team are specialized and employed by different institutions (ACDHS, UPMC and AHN) by design; however, this also means that they work with different data systems and are familiar with different methods of viewing data results. The LLM summaries support the multidisciplinary team's case conferencing process and provide a common foundation for decision-making. They are particularly useful for understanding a client's broad service history in the context of SST, in which eligible clients have co-occurring conditions.

SST team members estimate that summaries save at least 30 minutes per client by providing a centralized record of the client's history of interactions with ACDHS and other County institutions. Members of the care team can use this information to dig deeper into EHR systems, County records and/or specific services. They can also use the information in the summaries to identify other providers who know the client. Connecting with these partners early supports better coordination and client service.

The ACDHS data science team trains SST team members on the use of the administrative data summaries, instructing them to use summaries as a tool to augment case conceptualization; they are also trained on the limitations of the data ingested by these summaries.

Summary Report Sections

- **Identifiers:** Reports for use by program staff include client first name, last name and date of birth. All data are pulled directly from the data warehouse; no modeling is used.
- **Administrative Data Summary:** The summaries follow a standard template. They describe the first event observed in a given domain, list the events that occurred in the last 12 months and describe any ongoing interactions with County programs. The domains included in the summaries are:
  - **Housing Services & Evictions:** This category describes interactions with County housing services. This may include shelter programs, street outreach and/or housing-first programs. The category also includes court records for eviction filings. The team uses this information for context on a client's current housing situation.
  - **SUD Services & Diagnoses:** This category includes services for treating SUD and chemical dependencies. The data are sourced from Medicaid claims and claims for County-paid services. Records of past services for substance use help care teams understand the acuity of a client's situation as well as their capacity to maintain engagement with a service provider.
  - **Behavioral Health Services & Diagnoses:** This category includes inpatient and outpatient services for treating behavioral health needs. The data are sourced from Medicaid claims and claims for County-paid services. The data also include records of engagement with ACDHS services. Behavioral health data are used to assess the extent to which a client fits the case conceptualization for SST.
  - **Criminal-Legal System Involvement:** This category includes court records from the Allegheny County court system and information from ACJ. Court data fall into broad categories of criminal charges (e.g., drug charges, traffic charges). ACJ data are limited to the start and end dates for periods of incarceration. If a person is transferred or released, their end date is presented in the same way. This field is used by care teams to understand if a client is currently incarcerated.
  - **Physical Health:** This category includes inpatient and outpatient services for treating physical health needs. The data are sourced from Medicaid claims and claims for County-paid services. They also include records from emergent care in regional hospitals, irrespective of payer type. Due to high rates of morbidity and mortality in the unhoused population, this information is useful for understanding clients' physical health needs.

    - **Children, Youth and Families (CYF):** This category includes data on referrals and home removals through the County's CYF programs. Some SST-eligible clients had significant interactions with CYF as children. This field can be used to understand a client's history as an adult or child but, because it provides just the date of the first event and a summary of the last 12 months, it offers intentionally limited visibility into a client's experience before the age of 18.
- **Recent Events Table:** The summary includes a table displaying a three-month history of interactions with County services using same categories listed in the Administrative Data Summary. This layout is sometimes more practical for understanding where a person was on a given date.

How We Build the Summaries

The construction of a summary is effectively a data pipeline with a single step that passes a prompt to an LLM. There is no multi-turn element of the design and the LLM is highly constrained. Significant work is done before LLM processing: extracting relevant information from the data warehouse, formatting it into a consistent schema optimized for LLM processing and passing it in with the prompt. When composing the Administrative Data Summary, relevant data are passed to the model in a stable, JSON-like input structure along with domain-specific instructions and examples. The prompt and examples explicitly discourage making inferences or recommendations on the basis of the data. Model responses are rendered into structured HTML templates which are shared as summaries of client-level administrative data for SST case conferencing. The only parties with access to the summaries are the SST care team and the team responsible for creating the summaries.

The Model

The system uses Qwen3-4B-Instruct, chosen for its small size and strong performance on tightly constrained summarization tasks. We intentionally avoided fine-tuning because the task is narrow, patterned and repetitive, and the combination of high-quality prompts with consistently structured inputs already produces stable and reliable results. Using a small model also keeps computing costs low, which is important given that the reporting workflow is run regularly to keep summaries current.

Validation

The goal of the LLM is to format structured data to be more legible and easily accessible to a member of the care team. The LLM, in formatting the data, should not make choices about what information to include or not include. For this reason, we view evaluation metrics that focus on writing quality (e.g., Fluency) or overlapping sentences (e.g., ROUGE 1) as less relevant than measures that explicitly check for cases where information from the input data is omitted or modified. We used an LLM-as-judge approach with a private GPT 4.1 deployment in our cloud tenant to evaluate performance on a sample of 5,000 summaries. Additionally, we used human review to assess the quality of the LLM judge.

We employed the following LLM-as-judge metrics:

- **Recommendations:** The fraction of summaries flagged by a GPT 4.1 evaluator to contain a recommendation or call to action for the reader
- **Information Modified:** The fraction of summaries flagged by a GPT 4.1 evaluator to contain event descriptions that are rewritten in a way that changes the meaning of an event from the input artifact
- **Information Excluded:** The fraction of summaries flagged by a GPT 4.1 evaluator that omitted an event in the input artifact

The final metric is event miscounts. We define this as the fraction of summaries featuring event counts that do not correspond to the count in the input artifact. Event miscounts were computed using regex.

Table 8 displays LLM performance based on these metrics.

TABLE 8: **Language Model Performance**

METRIC	QWEN3-4B-INSTRUCT	QWEN2.5-32B	LLAMA3.3-70B
LLM-as-judge			
Recommendations	0.00%	0.00%	0.00%
Information Modified	1.12%	2.80%	3.44%
Information Excluded	0.04%	0.12%	0.12%
Event Miscounts	0.00%	0.00%	0.00%
n clients	777	397	397
n summaries	5,000	2,500	2,500

Based on the evaluation metrics above, we selected Qwen3-4B-Instruct for the program pilot and are exploring opportunities to further improve performance. We note that larger models exclude and modify information at marginally higher rates than Qwen3-4B-Instruct and attribute the difference to the extent of changes introduced that appear to make the descriptions more idiomatic. During the pilot, we are explicitly guiding the small number of users to check the information in summaries and use other sources for their final case conceptualization.

In order to check the performance of the LLM-as-judge process, we sampled 240 summaries from the original 5,000, while oversampling cases in which the LLM-as-judge process flagged errors (**Table 9**).

A human review of LLM-as-judge evaluations disagreed with 0.8% of information-modified labels. The majority of false positives we identify are cases in which the models flag modifications that change the phrasing but do not alter the content of the input. Given that they do not change the meaning of the summaries or introduce misleading information, we view these changes as false positives for the purpose of our evaluation. Conversely, the true positives and false negatives we identify are attributable to the summary simplifying event descriptions to a degree that loses content.

TABLE 9: **Manual Evaluation of LLM-as-Judge Labels, n=240**

		HUMAN FLAGGED	HUMAN DID NOT FLAG
Recommendations	Model Flagged	0	0
	Model Did Not Flag	0	240
Information Modified	Model Flagged	61	7
	Model Did Not Flag	12	160
Information Excluded	Model Flagged	1	1
	Model Did Not Flag	0	238
Event Miscount	Model Flagged	0	0
	Model Did Not Flag	0	240

Safeguards

In addition to our extensive testing, we also put into place three important safeguards designed to minimize the likelihood of errors that may impact an individual's assessment:

- Members of the care team always have access to the raw data inputted into the LLM. While those data are slower to parse, the ability to view them provides an important safeguard against LLM hallucination.
- Issues with the model will be reported immediately to the ACDHS data science team for investigation.
- On an ongoing basis, the data science team is evaluating how the availability of an LLM summary impacts care team behaviors.

The data science team continues to iterate towards best practices regarding the use of LLM technology, including coordinating with experts to give advice on the optimal ways to process the data and share the results.

AUTHORSHIP

Authors

Nic Marlton, Anne Kainaroi and Pim Welle, PhD, Allegheny County Department of Human Services

Reviewers

Katherine Rittenhouse, PhD (University of Texas at Austin), Holly Wiberg, PhD (Carnegie Mellon University) and Jessica M. Gannon, MD (University of Pittsburgh)[1]

Acute Working Group Members

The following were the initial members of the Acute Working Group and the co-designers of the SST intervention: Todd Wahrenberger, MD, Madeliene Lipshe Williams, MD, Jodie Bryk, MD, Max Hurwitz, MD, Divya Venkat, MD, Heidi Allison, Jamie Gribben, Justin Grotelueschen, Stuart Fisk, Anne Kainaroi, Pim Welle, PhD, Nic Marlton

Supporting Stakeholders

The following people were key contributors and led the way to the realization of SST at ACDHS:
Jenn Batterton, Katy Collins, Sarah Dansforth, Alex Jutca, Madeleine Lipshe Williams, MD

The following people were key contributors in developing, reviewing and implementing the data-driven models:
Erika Montana, Caleb Rollins, Steven Lann, Ethan Goode, Abhishek Jallawaram, Honey Rosenbloom

Additionally, Maisie Taibbi, Ben Talik, Brian Wolford and Daniel Auflick provided subject matter expertise.

---

1 Did not review **Methodology** or **Appendix B**

APPENDIX A: SUMMARY OF STAKEHOLDER ENGAGEMENT MEETINGS

Stakeholder Engagement Meetings

Beginning in late 2023, ACDHS gathered input from stakeholders and practitioners in community mental health and homelessness services to inform the development of and refine the SST Design. **Table A1** outlines the purpose and timing of each event.

TABLE A1: **Stakeholder Engagement Meeting History**

MEETING	DESCRIPTION	DATE
Homelessness Services Stakeholder Meetings	Initial meetings with stakeholders from UPMC, AHN, the City of Pittsburgh and ACDHS Homelessness Services to acknowledge gaps in existing services and discuss options for serving clients with sever and unmet behavioral health needs.	10/2/2023 12/11/2023 4/4/2024
Acute Presentations Working Group Meetings	Regular meetings focused on designing a concrete intervention for supporting unhoused County residents with sever and unmet behavioral health needs. These meetings included detailed review of potential data-driven referral methodologies. Working group members were enlisted as advisors on the design of the composite methodology.	5/9/2024 5/16/2024 5/23/2024 5/30/2024 6/6/2024
Downtown Case Conferencing	Regular case conferencing for individuals with sever and unmet behavioral health needs in downtown Pittsburgh, attended by UPMC, AHN and County representatives. These case conferencing meetings informed the recommendations for and design of SST.	7/11/2024 7/18/2024 9/9/2024
Recommendations Presentation and Discussion	Shared recommendations from the working group and case conferencing process with stakeholders from UPMC, AHN, the City of Pittsburgh and ACDHS Homelessness Services. Participants in this discussion helped to further refine the program design.	10/7/2024
Opioid Settlement Fund Listening Sessions	Public listening sessions to gather feedback on allocation of opioid settlement funds. Community members and providers attended these sessions and advocated for design elements—like a focus on harm reduction—that are fundamental to SST.	10/8/2024 10/15/2024
Administrative Referral List Review with Outreach Workers	Meetings with frontline workers from the ACDHS Coordinated Entry Field Unit and Pittsburgh Mercy: Operation Safety Net to review list of possible data-driven referrals. Recommendations and feedback were used to refine the data-driven model.	10/7/2025 10/17/2025
Homeless Outreach Coordinating Committee	Presentation to homelessness services provider staff within the CoC, focused specifically on providing transparency on the design of SST—including the two referral pathways and data-driven model implementation—and soliciting feedback from homelessness service providers.	10/13/2025
Street Medicine Provider Committee	Presentation to a cohort of ~30 street medicine practitioners in Allegheny County, focused specifically on providing transparency on the design of SST—including the two referral pathways and data-driven model implementation—and soliciting feedback from homelessness service providers.	10/16/2025

MEETING	DESCRIPTION	DATE
Homelessness Services Stakeholder Meeting	Presentation to a cohort of ~40 healthcare practitioners representing both major hospital systems, focused specifically on providing transparency on the design of SST—including the two referral pathways and data-driven model implementation—and soliciting feedback from homelessness service providers.	12/12/2025
Administrative Referral Reviews with Outreach Workers	Reviewed the pre-launch list of 100 data-driven referrals with frontline workers from Pittsburgh Mercy: Operation Safety Net, the ACDHS Community Services Field Unit and the Office of Community Health & Safety: ROOTS. Input from this group was tracked in case notes and used for case conferencing prior to the program launch.	4/1/2026 4/8/2026 4/14/2026 4/15/2026
Street Medicine Provider Committee	Presentation to a cohort of ~30 street medicine practitioners in Allegheny County, focused specifically on reminding participants about the program launch date and the process for submitting provider referrals.	4/16/2026
Allegheny County Field Unit Huddle	Presentation to frontline workers in the outreach arm of the County's coordinated entry system, focused specifically on reminding participants about the program launch date and the process for submitting provider referrals.	4/23/3026
Homelessness Outreach Coordinating Committee	Presentation to homelessness services provider staff within the CoC, focused specifically on reminding participants about the program launch date and the process for submitting provider referrals.	5/11/2026
Onsite Pre-Launch Coordination with Frontline Workers	Half-day onsite visits with frontline workers from UPMC Resolve, Downtown Court, Second Avenue Commons and the Office of Community Health & Safety: ROOTS to coordinate service delivery during the launch of SST.	4/15/2026 4/20/2026 5/18/2026 5/26/2026

APPENDIX B: DATA-DRIVEN MODEL IMPLEMENTATION DETAILS

Outcomes

The data-driven model used for SST referrals is a composite of four component models. **Table A2** provides more detail about the outcomes used to train the component models.

TABLE A2: **Data-Driven Model Predicted Outcome Detail**

OUTCOME NAME	DESCRIPTION	BASE RATE IN COUNTY	REASON FOR INCLUSION
Involuntary hospitalization (302) within one year	A binary outcome (0,1) equal to one if there is an upheld 302 petition for an individual in the year following the referral date.	2,509 0.47%	Legally, 302s in Allegheny County are designed to allow for intervention when individuals managing mental health crises present a danger to themselves or others. For SST referrals, it is used as a proxy for individuals for whom step-down behavioral health services are not working.
ACJ booking within one year	A binary outcome (0,1) equal to one if an individual is booked in ACJ in the year following the referral date.	5,870 1.10%	The cohort for whom SST was designed has a materially higher rate of interaction with first responders than the general population. For SST referrals, ACJ booking is used as a proxy for when an individual has a negative interaction with the criminal-legal system.
Indication of being unhoused within one year	A binary outcome (0,1) equal to one if a person checks into a County shelter, continues to stay in a shelter for at least 60 days or interacts with a street outreach team in the year after the referral date.	2,515 0.47%	SST is designed to serve people who are unhoused. For SST referrals, encounters with a street outreach team or staying in a shelter is a proxy for a person experiencing homelessness in the future.
Fatal overdose within one year	A binary outcome (0,1) equal to one if there is a Medical Examiner determination of death by overdose on a date within one year after the referral date.	338 0.06%	SST serves people who demonstrate profound disability and/or high vulnerability and are chronically disengaged from clinic-based care due of the severity of psychiatric symptoms inclusive of substance use disorder (SUD). For SST referrals, a Medical Examiner determination of fatal overdose is used to identify individuals experiencing SUD. Given the rarity of this outcome, we are exploring opportunities to replace or augment it as data become available.

Model Architecture

The data-driven models, as implemented, are gradient-boosted ensembles of decision trees, using the XGBoost library. Boosted trees can capture complex patterns while protecting from overfitting through structural hyperparameters. We chose a subset of the selected hyperparameters to describe the general structure of each model:

- **Max depth:** An integer value controlling how deeply a tree can capture interactions between features
- **Subsample:** A decimal value (0, 1] controlling the fraction of rows randomly sampled for each tree
- **Column sample by tree:** A decimal value (0, 1] controlling the fraction of columns randomly sampled for each tree
- **Number of estimators:** An integer value controlling how many boosting rounds the model performs

TABLE A3: **Data-Driven Hyperparameters**

HYPERPARAMETER	ACJ	OVERDOSE	302	SHELTER
Max depth	95	20	87	95
Subsample	96.5%	88.3%	83.7%	98.3%
Column sample by tree	52.9%	55.0%	52.4%	60.6%
Number of estimators	941	880	1179	382

Features

Table A4 shows the counts of features used in the data-driven models, organized by domain. We intentionally excluded information older than five years to uphold an individual's right to be forgotten and did not observe any significant decrease in performance by limiting the time horizon in this way. We selected features using Lasso. Features in the model may take the form of count, duration, number of days since or a binary flag indicating if service use is ongoing. To deal with missingness in our source data, we zero impute and include a missing flag for features which have no value for an individual (e.g., for transformations that require a value to be meaningful, as in the case of the count of days since an event). Interactions between features were not modeled explicitly but are included implicitly given that the models were implemented using a gradient-boosted tree architecture.

TABLE A4: **Data-Driven Model Features by Domain**

CATEGORY	DESCRIPTION	302	ACJ	SHELTER	OVERDOSE
Basic Needs Assistance	Requests for basic needs from 2-1-1	0	1	15	0
CYF	CYF referrals, investigations and cases	0	13	2	0
Criminal-Legal System	Counts of charge types, previous length of stay in ACJ, bail amount	9	148	71	22
Demographics	Race, age, gender, education level	20	20	20	2
Family	Recent birth events	0	1	1	0
Housing	Shelter stays, street outreach	3	23	46	10
Mental Health	Mental health diagnoses and services	56	41	40	13
Physical Health	Emergency and inpatient physical health hospital stays	12	23	29	24
Public Benefits	Medicaid, Supplemental Security Income (SSI), Temporary Assistance to Needy Families (TANF)	0	3	1	4
SUD	Diagnoses, substance use treatment, outpatient visits	4	53	45	41
Total		**104**	**326**	**270**	**116**

Training Data Demographics

Table A5 outlines basic demographic information (from the Allegheny County data warehouse) for the people included in the training data used in the data-driven model. The as-of date for the training data was 6/10/2024. On this date, a meaningful fraction of those in the data set had missing or incomplete demographic data, accounting for the six percent of individuals with unknown values in description of race.

TABLE A5: Data-Driven Model Training Data Demographics Detail

CATEGORY	GROUP	COUNT	%
Race Description	White	335,184	63.1%
	Black	127,531	24.0%
	Other	35,166	6.6%
	Unknown as of 6/10/2024	33,388	6.3%
Legal Sex Description	Male	248,573	46.8%
	Female	236,557	44.5%
	Unknown as of 6/10/2024	46,139	8.7%
Age	18 through 34	177,292	33.4%
	35 through 54	188,689	35.5%
	55 through 64	74,491	14.0%
	65+	74,698	14.1%
	Unknown as of 6/10/2024	16,099	3.0%
n	-	531,269	-

Model Evaluation Metrics

We evaluated the model's performance using multiple metrics to assess different aspects of predictive accuracy. **Table A6** outlines the metrics used and their interpretation.

TABLE A6: Data-Driven Model Evaluation Dictionary

METRIC	INTERPRETATION
AUROC	The likelihood that someone who will experience the outcome receives a higher score than someone who will not
Precision	Percent of eligible individuals who experience the predicted outcome within the next 12 months
Recall	Percent of eligible people who experience homelessness within the next six months
Calibration Slope	Close to 1 means a predicted risk can be interpreted as a probability
Calibration Intercept	Close to 0 means a predicted risk can be interpreted as a probability

Out-of-Sample Performance and Calibration

In general, we leave interpretation of the model performance metrics to the reader.

One exception is the calibration slope for the model predicting fatal overdose. This metric tells us the extent to which predicted probabilities from the model match observed outcome rates. Notably, the model that predicts fatal overdose in the next 12 months had significantly worse calibration than the other three. We attribute this calibration miss to the number of observations in our training data (only 338).

Based on its slope, it is fair to say that the model predicting overdose is overestimating fatal overdose risk. Steps taken to reduce its impact on the program are described in the composite methodology section of **Appendix B**.

TABLE A7: Out-of-Sample Performance for Data-Driven Models

METRIC	302	JAIL	SHELTER	OVERDOSE
Precision (Top 1,000)	34.7%	62.3%	70.1%	3.3%
Index to base rate (Top 1,000)	73.5X	56.4X	148.1X	51.9X
Recall (Top 1,000)	13.8%	10.6%	27.9%	9.8%
Precision (Top 1%)	16.2%	40.4%	25.9%	2.1%
Recall (Top 1%)	34.3%	36.6%	54.7%	33.4%
Precision (Top 5%)	5.7%	16.6%	7.2%	0.8%
Recall (Top 5%)	60.4%	75.1%	76.0%	64.5%
AUROC	0.8761	0.9522	0.9454	0.9009
Calibration Intercept	0.0001	0.0002	0.0001	0.0001
Calibration Slope	0.9934	0.9994	1.0017	0.8311

Performance Comparisons Among Demographic Subgroups by Component Model

Tables A8-A11 detail the extent to which precision, recall and calibration differed by subgroup for the 1000 people with the highest assessments according to the data-driven model. We leave interpretation to the reader.

TABLE A8: Performance by Subgroup for the Involuntary Hospitalization (302) Model

POPULATION	N (TOP 1,000)	PRECISION	RECALL
All	1,000	34.7%	13.8%
Male	508	32.5%	13.5%
Female	463	35.9%	14.8%
White	524	33.6%	11.5%
Non-White	468	35.7%	18.0%
Past 302	968	34.7%	37.6%
No Past 302	32	34.4%	0.7%
Past Homelessness	259	34.0%	34.8%
No Past Homelessness	741	35.0%	11.5%
Past ACJ	333	37.2%	26.7%
No Past ACJ	667	33.4%	10.9%

TABLE A9: Performance by Subgroup for the ACJ Booking Model

POPULATION	N (TOP 1,000)	PRECISION	RECALL
All	1,000	62.3%	10.6%
Male	705	62.3%	10.9%
Female	226	65.0%	11.0%
White	431	62.2%	10.9%
Non-White	566	62.5%	10.5%
Past 302	185	67.0%	19.9%
No Past 302	815	61.2%	9.5%
Past Homelessness	372	65.1%	28.3%
No Past Homelessness	628	60.7%	7.6%
Past ACJ	982	62.3%	17.9%
No Past ACJ	18	61.1%	0.4%

TABLE A10: Performance by Subgroup for the Interaction with Shelter Services Model

POPULATION	N (TOP 1,000)	PRECISION	RECALL
All	1,000	70.1%	27.9%
Male	624	70.0%	30.0%
Female	355	69.9%	26.4%
White	485	73.8%	32.1%
Non-White	495	66.7%	24.6%
Past 302	119	71.4%	30.7%
No Past 302	881	69.9%	27.5%
Past Homelessness	1000	70.1%	52.2%
No Past Homelessness	0	-	-
Past ACJ	399	69.2%	31.8%
No Past ACJ	601	70.7%	25.8%

TABLE A11: Performance by Subgroup for the Fatal Overdose Model

POPULATION	N (TOP 1,000)	PRECISION	RECALL
All	1,000	3.3%	9.8%
Male	479	4.4%	11.5%
Female	419	2.1%	7.4%
White	826	3.0%	10.0%
Non-White	174	4.6%	9.2%
Past 302	142	4.2%	21.4%
No Past 302	858	3.1%	8.7%
Past Homelessness	213	3.3%	25.9%
No Past Homelessness	787	3.3%	8.4%
Past ACJ	509	3.7%	20.4%
No Past ACJ	491	2.9%	5.7%

Composite Model Methodology

We designed the SST composite model to identify a cohort of unhoused adults who demonstrate profound disability and/or high vulnerability and are chronically disengaged from clinic-based care due to the severity of psychiatric symptoms including but not limited to SMI and SUD. We took two complementary approaches to creating this composite. First, we attempted to maximize the rate at which one or more of the predicted outcomes was observed in the people with the highest assessments. Then, using a composite methodology, we reviewed lists of people referred by practitioners and leaders in homelessness services, street medicine, inpatient psychiatry and behavioral health crisis response. Through this process, we arrived at the set of models and filters for the SST composite.

An important question is how to blend outputs from multiple data-driven models. We cannot simply sum (or average) the outputs of the risk models because rare events (e.g., mortality) might be more severe than more common events (e.g., homelessness). To address this, we implemented an approach that involved projecting individual model's predictions into a standard normal distribution (z-scores) before summing their values. In this way, we normalized the probabilities so that a single unit in standard normal space counted the same across outcomes.

More sophisticated weighting methodologies might attempt to trade the value of outcomes against each other using the Value of Statistical Life (VSL) and Quality Adjusted Life Year (QALY) / Disability Adjusted Life Year (DALY) frameworks. We leave that for future work, noting that our existing compositing methodology is highly effective at identifying individuals at risk across each of our outcomes of interest and thus functionally serves its purpose.

The formula for generating composites is as follows:

$$Composite = \sum_{i=1}^{4} P(F(M(\hat{y}_i))) * w_i$$

Where:

- ***P(x)*** is the inverse cumulative distribution function of the standard normal distribution:

$$P: (0,1) \to \mathbb{R}$$

- w_i is defined as:

$$w_i = \begin{cases} {}^1/_4, for\ the\ fatal\ overdose\ model \\ 1, for\ all\ other\ models \end{cases}$$

- ***F(x)*** is defined as:

$$F(x) = \frac{x}{2000}$$

- ***M(x)*** maps a real value into one of 1999 quantile bins, returning an integer:

$$M: \mathbb{R} \to \{1, \dots, 1999\}$$

Given the small size of the outcome data for the overdose model, we made an explicit adjustment to reduce its overall impact on the referral process. Readers can see this adjustment in the piece-wise definition of w_i.

Composite Model Performance

Table A12 shows performance metrics by demographic subgroup for the composite model among the 1,200 individual referrals used for evaluation. We calculated the metrics by evaluating the fraction of individuals for whom one or more outcomes occurred. Readers may notice that this table excludes the section comparing individuals with and without past homelessness. This exclusion was based on the fact that SST data-driven referral assessments are only conducted for people who have a record of past homelessness in the Allegheny County data warehouse.

TABLE A12: **Performance by Subgroup for SST Data-Driven Model Referrals**

POPULATION	N REFERRALS	PRECISION	RECALL
All	1,200	89.8%	8.7%
Male	831	88.1%	8.8%
Female	369	93.8%	8.5%
White	625	88.3%	9.4%
Non-White	575	91.5%	8.1%
Past 302	693	88.3%	34.9%
No Past 302	507	91.9%	4.4%
Past ACJ	1132	90.1%	20.0%
No Past ACJ	68	85.3%	0.8%

Comparison Heuristic Methodology

To benchmark the performance of a data-driven model referral system, we used a heuristic that approximates the case conceptualization for SST and compared it to our composite methodology. The cohort had at least one record in the Allegheny County data warehouse for each domain.

TABLE A13: **Heuristic Selection Methodology**

DOMAIN	FIELDS USED	REASONING
Criminal-Legal System Involvement	ACJ booking in the prior year.	The SST cohort faces elevated risk of criminal-legal system involvement.
Substance Use-Related Medical Claim	Medical claim related to an Alcohol Use Disorder (AUD), Opioid Use Disorder (OUD) or SUD diagnosis in the prior year.	The SST cohort is navigating substance use challenges.
Behavioral Health Diagnosis	Medical claim in the last year related to bipolar disorder, personality disorder, a condition involving delusions or psychosis other than schizophrenia, schizophrenia or schizoaffective disorder and/or other adult onset mental health disorders.	The SST cohort case conceptualization includes severe behavioral health issues. We included a broad set of behavioral health diagnoses to ensure that heuristic filtering conditions were not too restrictive.
Evidence of Homelessness	Interaction with an entry-exit shelter, night-by-night shelter and/or street outreach team in the last year.	Individuals who have not experienced homelessness in the last year are ineligible for SST.