RISK STRATIFICATION, SEGMENTATION, AND TIERING TRANSPARENCY DOCUMENT

SUMMARY

DECEMBER 2025

DHCS

CALIFORNIA DEPARTMENT OF
HEALTH CARE SERVICES

Table of Contents

RSST Transparency Summary

Introduction

This summary distills the key components and intent of the original *RSST Transparency Document (July 2025)* published by the California Department of Health Care Services (DHCS). It provides an overview of how RSST works, what makes it unique, and why it represents a major step forward in proactive, data-driven, and equitable population health management across the state. The document is intended to be accessible to readers of all statistical backgrounds.

RSST: California's Data Analytics Solution

California's Medi-Cal program is implementing a statewide algorithm called Risk Stratification, Segmentation, and Tiering (RSST) to better identify members who are at risk of poor health outcomes or not engaging in needed care. This algorithm is part of a broader Population Health Management (PHM) strategy aimed at improving equity, care coordination, and outcomes across the state.

RSST serves as a central component of Medi-Cal Connect, a statewide data analytics solution designed to support population health management. The primary goal of RSST is to enable care teams and MCPs to proactively identify and support members who are at risk *before* health problems escalate. RSST's predictive capabilities help MCPs allocate resources more effectively, while its focus on equity ensures that unmet or overlooked needs—such as missed preventive visits, unmanaged behavioral health conditions, or housing instability—among higher-risk populations are identified. This allows for targeted outreach and care coordination where support will have the greatest impact.

The system is built on four core principles:

- » **Standardization** across all Managed Care Plans (MCPs) to ensure consistency.
- » **Transparency** in how models are built and used.
- » **Equity**, with a focus on identifying and reducing disparities.
- » **Flexibility** to adapt to future policy changes without needing a full redesign.

Why RSST?

DHCS developed RSST to address current challenges in risk identification throughout the Medi-Cal delivery system, including inconsistent stratification methods across MCPs,

reliance on fragmented or incomplete data, and a narrow focus on cost or utilization. Many existing models are proprietary and lack transparency, offer little attention to equity, and/or generate risk scores based on past events rather than future risk. These tools often treat all patient groups uniformly, without accounting for the diverse needs across populations, and are updated infrequently, limiting their responsiveness to changing circumstances.

RSST directly addresses these gaps by introducing a statewide, standardized method that applies to all Medi-Cal members, ensuring consistency and fairness across MCPs. The algorithm draws on centralized, comprehensive data from the Medi-Cal program, including sources not readily accessible to individual MCPs, such as behavioral health conditions and enhanced member demographic and contact information.

Predictive Risk Tiering

The RSST algorithm uses advanced machine learning to analyze a wide range of Medi-Cal data—including claims, encounters, demographic information, enrollment history, and social drivers of health. RSST examines three risk domains for both adults and children—adverse health events, underutilization, and social risk—and each domain has associated subdomains (see Table 1). For a complete list of outcomes, see Appendix A.

Each month and based on data availability, the algorithm predicts the likelihood that a member will experience negative health events or underuse essential health care within the next year and assigns every member to one of three risk tiers: low, medium-rising, or high. Risk tiers are defined as follows:

» **Low Risk**: Stable health, low predicted need for intervention

» **Medium-Rising Risk**: Emerging or moderate needs, often a focus for preventive outreach

» **High Risk**: High predicted need or likelihood of adverse outcomes; requires further engagement.

If a member is high-risk in any single subdomain, they are categorized as high-risk overall, ensuring that significant needs are not overlooked, even if they are concentrated in a single area. Member tiers are generated monthly based on updated data, ensuring that care teams have up-to-date insights to guide interventions.

Table 1. RSST Algorithm Structure

Domain Risk Tier	Risk of Adverse Events		Risk of Underutilization[1]		Social Risk
Subdomain Risk Tier	**Physical Health**	**Behavioral Health**	**Physical Health**	**Behavioral Health**	**Social Events**
Example Outcomes (Adult Model)[2]	» Inpatient admission » ED visits » New chronic condition diagnosis	» New diagnosis of mental illness » Psychiatric admission » Psychiatric ED visit	» Well child visits » Primary care visits » Dental care visits » Medical or pharmacy claims	» Ambulatory MH or SUD visits » PCP Visits with BH or SUD diagnosis	» Housing Instability

[1] *Underutilization* refers to members who, based on their health conditions and recent interactions with the Medi-Cal system, should have received certain follow-up or preventive service—but did not.

[2] For a complete list of outcomes, see Appendix A.

Tiering Methodology

To select the most equitable and effective strategy for defining high-risk thresholds, DHCS evaluated three different tiering options:

1. **Equal Proportion Approach:** Flagging the same percentage of members as high-risk in each subdomain.

2. **Balanced Recall Approach:** Setting thresholds to ensure similar sensitivity (recall) for high-risk designation across all subdomains, thus capturing a consistent share of true outcomes in each area.

3. **Maximized Recall Approach:** Optimizing thresholds to maximize overall recall, favoring subdomains with stronger predictive performance.

Each strategy was assessed for technical performance, balance across risk types, equity in subgroup recall, potential benefits (e.g., identifying "net new" at-risk members), and operational impacts. Ultimately, DHCS chose Option 2 to balance recall (sensitivity)

across all subdomains. This approach ensures that members at risk in any domain are identified consistently, regardless of the specific outcome or population subgroup.

High-Risk Ceiling

Beginning no sooner than July 2026, MCPs will be required to assess members identified by RSST as high risk. To support implementation, DHCS established an initial high-risk tier ceiling so that no more than 10% of Medi-Cal members across the state are classified as "high risk" This ceiling is informed by policy guidance and stakeholder input and reflects what is operationally feasible for MCPs to assess and engage. This is then cascaded to each responsible MCP (note that the 10% limit is not applied at the MCP level, which may result in MCPs with more or less than 10% of each MCPs membership appearing as high risk). In the future, DHCS will adjust the 10% high-risk ceiling based on what is best for members and what MCPs can support operationally.

Distinctive Features

Dual Focus on Risk and Underutilization: Unlike most traditional models that only flag high-cost or high-need individuals, RSST uniquely identifies members who are not using services they likely need—a sign of hidden or unmet needs for health care and social services. RSST's dual focus on risk and underutilization is essential for surfacing gaps in care that might otherwise remain invisible, supporting proactive and equitable intervention.

RSST's underutilization measures identify members who, based on their health conditions and recent interactions with Medi-Cal, should have received certain follow-up or preventive services but did not. For these measures, the denominator includes members who have recently engaged with the health care system and have a diagnosis that requires an outpatient follow-up within the next 12 months. It also includes members who are likely to require dental care or pharmacotherapy for conditions that would typically warrant follow-up, along with members with no medical or pharmacy claims. In other words, these are individuals for whom a follow-up is clinically indicated. The numerator is the subset of those members who did not receive the needed follow-up in the subsequent 12 months. This approach helps pinpoint where gaps in care are occurring and highlights potential barriers to access.

This RSST approach advances DHCS's ability to surface many members previously not recognized as underutilizing the services they likely need.

Behavioral Health and Social Risk: The inclusion of behavioral health and social risk factors makes RSST unique because it goes beyond traditional models that focus solely on physical health or health care utilization. By integrating predictors such as mental health conditions, substance use, housing instability, and other social drivers, RSST provides a more holistic and accurate picture of member risk.

Max-Tier Logic: If a member is high-risk in any single subdomain, they are classified as high-risk overall. This "max tier" approach ensures that significant needs are not overlooked, even if needs are concentrated in only one area.

By not averaging or diluting risk across subdomains, RSST prevents significant concerns from being masked by lower risk in other areas, ensuring that members with urgent or complex challenges receive timely outreach and care coordination. This method is especially important for surfacing specific but critical needs that might otherwise be missed in a more generalized risk assessment.

Standardization and Equity: RSST uses the same methodology statewide, across all members, reducing variability and supporting equitable care. Equity and monitoring are built into every stage, from data selection and model development to validation and performance tracking.

Technical Design

RSST is composed of ten predictive models—five for adults and five for children—each trained on nine years of historical Medi-Cal data that includes professional, facility, dental, behavioral health, and pharmacy claims, as well as detailed member enrollment information for over 15 million individuals. These models draw from two primary data sources: Medi-Cal administrative claims and eligibility data, both standardized by DHCS using the All-Payer Claims Data Common Data Layout (APCD-CDL).

RSST incorporates a wide array of information such as prior hospitalizations, missed appointments, chronic conditions, social factors, and changes in coverage or residence. The predictive framework utilizes 31 variables specific to underutilization and 34 variables related to adverse health events, allowing for a nuanced and holistic evaluation of member risk.

The training process uses sophisticated data partitioning and "what was known when" logic to accurately reflect real-world prediction scenarios. The models predict a variety of binary outcomes for each member over the next 12-months using advanced but

proven machine learning techniques widely adopted in health services research and public health—primarily XGBoost (gradient-boosted decision trees), with LightGBM and regularized logistic regression as comparators.

Model development took place in a distributed, GPU-accelerated computing environment, enabling efficient processing of large-scale data. Ongoing performance is closely monitored through internal dashboards, with regular retraining and refinement to ensure the models remain accurate and equitable as population health changes. For complete details about technical design, refer to the *RSST Transparency Document (July 2025)*, beginning on page 10.

Ensuring Equity and Model Performance

Statistical validation tests demonstrate that RSST is highly effective at predicting both future health events and underutilization of care, successfully identifying members whose needs might otherwise go unnoticed. The system is calibrated so that the high-risk tier includes approximately ten percent of the Medi-Cal population, striking a balance between effective resource targeting and health plan capacity. RSST's technical performance is robust, with area under the curve (AUC)[1] scores ranging from 0.75 to 0.94 across models. Recall[2] for the high-risk tier is approximately 0.23–0.24, and precision metrics show that more than half of those flagged as high-risk subsequently experience an adverse outcome—demonstrating both accuracy and practical value.

Equity has been a core focus throughout RSST's development. To ensure the model performs equitably, its predictive accuracy is rigorously evaluated across key subgroups—including race, ethnicity, language, sex, and new Medi-Cal enrollees. Specifically, we assess Recall, which measures the model's ability to correctly identify all members who truly need support (i.e. true positives). The goal is to maintain subgroup recall rates within 80% of the statewide average, helping ensure that no group is systematically under-identified. This approach supports equitable identification of high-risk members across the diverse Medi-Cal population. DHCS is committed to ongoing monitoring, using industry-standard metrics such as AUC, recall, and calibration, and will

---

[1] *AUC* indicates how well models can predict people who will experience an outcome versus people who will not. An AUC of 1.0 means a model predicts outcomes perfectly; 0.5 means the model is no better than flipping a coin.

[2] *Recall* (i.e., sensitivity) measures how many of the true cases of an outcome the model successfully identifies.

regularly retrain and validate the models to ensure they remain accurate and equitable as populations and care patterns evolve.

Collaborative Development

The RSST algorithm was developed through a rigorous and collaborative process. It was guided by a robust leadership and governance structure to ensure methodological rigor, stakeholder alignment, and real-world relevance.

DHCS established dedicated internal leadership and the RSST Work Group to oversee strategy and execution. An Academic Work Group of national experts in health equity, Medicaid analytics, and machine learning provided technical guidance. Members brought deep expertise in ethical AI, advanced statistical methods, and causal inference—contributing cutting edge insights from their work. Their expertise helped ensure the model was grounded in both methodological innovation and real-world applicability. A Scientific Advisory Council—including representatives from MCPs, care delivery systems, and academic institutions—advised on key decisions and ensured the model's practical applicability.

Further, MCPs were actively involved through the Medi-Cal Connect Advisor process, where selected MCPs reviewed and validated RSST risk tier assignments against their internal assessments and care management activities. These MCPs provided feedback on the algorithm's accuracy, usability, and policy implications. This multi-level governance approach prioritized evidence-based methods, transparency, and stakeholder input at every stage of algorithm development.

Limitations

While RSST represents a significant advance, it's still in its early stages and it does have limitations. The system currently relies on claims and eligibility data resulting in data lags and does not yet include broader clinical sources like electronic health records. Claims data can also be incomplete, which may affect precision when identifying at-risk members. Certain member groups, such as Dual Eligibles (i.e., members who are eligible for both Medicaid and Medicare), are excluded due to incomplete data. This creates blind spots for members receiving care elsewhere, though this represents a small portion of the overall population. As Medi-Cal's population evolves, tier thresholds and definitions may need to be updated.

Conclusion

RSST lays the foundation for a more proactive, data-driven, and equitable Medi-Cal system. By systematically identifying not only those at highest risk, but also those under-engaged with care, RSST will help DHCS and its Medi-Cal partners close gaps and improve outcomes for California's most vulnerable residents. The RSST algorithm is open-sourced, promoting transparency, collaboration, and continuous improvement across the health care ecosystem. Ongoing updates and stakeholder collaboration will ensure the tool continues to adapt and drive improvements in population health statewide.

To learn more, visit dhcs.ca.gov/medi-cal-connect.

Appendix A. RSST Outcomes

Domain	Subdomain	Adult Models	Pediatric Models
Adverse Events	Physical Health	» Morbidity » Mortality » All Cause Inpatient Admissions » 3+ All Cause ED Visit	» Morbidity » Mortality » New Diagnosis of Common Chronic Illness » 3+ All Cause ED Visit
	Behavioral Health	» Care for Unintentional Drug OD » Injection Drug-Related Adverse Event » Care for Intentional Self-Harm » Psychiatric Admission » 2+ Psychiatric ED Visits » Co-Occurring High ED Utilization and MH/SUD Care	» New Diagnosis of Mental Illness » New Diagnosis of Developmental Delay » New Diagnosis of SUD » Psychiatric Admission » Psychiatric ED Visit » Care for Drug Overdose » Care for Intentional Self-Harm
Underutilization	Physical Health	» Underuse of Primary Care Visits » Underuse of Dental Care » Underuse of STI Screening (Chlamydia) among Females 16-24 » Underuse of Appropriate Pharmacotherapy for Common Indications » Underuse of Asthma Controller Medications » No Medical or Pharmacy Claims	» Underuse of STI Screening (Chlamydia) among Females 16-24 » Underuse of Well-Child Visits » Children & Adolescents with No Claims Utilization » Underuse of Topical Fluoride and/or Dental Care » Underuse of Immunizations » Underuse of Asthma Controller Medications

Domain	Subdomain	Adult Models	Pediatric Models
Underutilization (continued)	Behavioral Health	» Underuse of MH/SUD Office Visits » Underuse of MH/SUD-Related Primary Care » Underuse of Antidepressant Medications » Underuse of Antipsychotics (Schizophrenia) » Underuse of Opioid Agonist Therapy (OUD)	» Underuse of PCP Visits (among members with BH/SUD) » Underuse of MH/SUD-Related Office Visits » Underuse of Metabolic Screenings (among members with antipsychotics) » Underuse of Follow-Up Visits » Underuse of ADHD Follow-Up Care
Social Risk	Adverse Social	» Housing Instability	» Housing Instability