



Effectiveness of the Alternative Responsive Pilot: Initial Outcomes

Policy Brief

REPORT HIGHLIGHTS:

- Senate Bill (SB) 21-118 **authorized the Alternative Response (AR) pilot** within Adult Protective Services (APS).
- SB21-118 **requires an independent evaluation** to build evidence for the AR practice and inform the pilot's future.
- This **policy brief** summarizes **pilot reach, implementation, and impact data**.
- A **quasi-experimental design** was used to generate initial causal evidence on the **effectiveness of AR** for the pilot period of January 4, 2023 through December 31, 2023 (n=2,345 unique AR Pilot cases).
- Findings illustrate the **AR Pilot is having a positive impact on at-risk adults** in Colorado by reducing re-entry and case length through collaborative engagement.

AUTHORS:

Courtney L. Everson, PhD
Sr. Project Director
Colorado Evaluation and Action Lab

Ernest Boffy-Ramirez, PhD
Sr. Researcher/Project Director
Colorado Evaluation and Action Lab

Erin Wickerham, MPH
Staff Researcher II
Colorado Evaluation and Action Lab

Abstract

Building Evidence for the Alternative Response (AR) Pilot

Senate Bill (SB) 21-118 authorized an Alternative Response (AR) pilot within the Colorado Department of Human Services, Adult Protective Services (APS). This pilot creates a dual-track model for tailoring APS response to risk level. Allegations of low-risk mistreatment and self-neglect are tracked to AR and higher-risk allegations are tracked to Traditional Response. The Colorado Evaluation and Action Lab at the University of Denver serves as the independent evaluator for the legislatively required outcomes study. The goal of the study is to understand the effectiveness of AR and inform the future of this practice model.

This interim findings report highlights select outcomes from implementation year one (January 4, 2023 through December 31, 2023). Results show the AR practice can improve collaborative engagement between APS staff and clients, which can help stabilize the client and improve well-being. Compared to equivalent cases in the pre-pilot period, repeat involvement in APS was significantly decreased by 4.71%. Case length was also reduced by 5.50 days. Clients with AR cases refused 1.30% less services compared to equivalent cases. Descriptive data show the AR Pilot is having a strong reach in pilot counties and is helping to support individuals with higher levels of social isolation and vulnerable conditions.

Initial results indicate the AR practice is a person-centered approach that can inform best practices for supporting at-risk adults, including a growing aging population. Findings from the full two-year outcomes evaluation will inform the future of the AR Pilot in Colorado and its potential for expansion statewide.

Table of Contents

Abstract	i
Table of Contents	ii
Acknowledgements	iii
Data Sources	iii
Suggested Citation	iii
Note on Gender-Inclusive Language	iii
Introduction	1
Participating Counties	1
Description of the Study	3
Evidence-Building Approach.....	3
Fidelity of Implementation.....	3
Descriptive Analysis.....	3
Quasi-Experimental Design	4
Key Findings	6
Fidelity of Implementation.....	6
Descriptive: Reach and Implementation	7
Inferential: Confirmatory Outcomes	12
Inferential: Exploratory Outcomes	16
Implications	17
Appendix A. Fidelity Measures	18
Appendix B: Quasi-Experimental Design Description and Matching Variables	20
Appendix C: Outcome Measures	22
Endnotes	23

Acknowledgements

This research was supported by the Colorado Department of Human Services (CDHS), Division of Aging and Adult Services. The opinions expressed are those of the authors and do not represent the views of the State of Colorado, CDHS, or the University of Denver. Policy and budget recommendations do not represent the budget or legislative agendas of state agencies, the Governor's Office, or other partners. Any requests for funding or statutory changes will be developed in collaboration with the Governor's Office and communicated to the legislature through the regular budget and legislative processes.

Data Sources

The study uses data from three sources:

1. Colorado Adult Protective Services (CAPS) administrative data system.
2. Fidelity of Implementation measures, using CAPS data.
3. Qualitative data collected through focus groups and pilot county feedback.

Suggested Citation

Everson, C.L., Boffy-Ramirez, E., & Wickerham, E. (Nov. 2024). *Effectiveness of the Alternative Responsive Pilot: Initial Outcomes (Policy Brief)*. (Report No. 21-09D). Denver, CO: Colorado Evaluation and Action Lab at the University of Denver.

Note on Gender-Inclusive Language

The Colorado Evaluation and Action Lab affirms our commitment to the use of gender-inclusive language. We are committed to honoring the unique gender identity of each study participant. Throughout this report, we follow the guidance of the Associated Press Stylebook and the Chicago Manual of Style and use the gender-neutral, singular “they” when appropriate.

Introduction

[Senate Bill \(SB\) 118](#) (Alternative Response Mistreatment At-risk Adults) passed in the 2021 legislative session, authorizing a pilot of the Alternative Response (AR) practice for responding to reports of low-risk mistreatment or self-neglect of an at-risk adult. Current law allows for only one type of response for a county department of human services, regardless of the risk level reported.

The AR Pilot enables a dual-track model to better tailor the response approach to the unique circumstances of the case and allegations. Track One is called Traditional Response (TR) and is reserved for higher-risk allegations of mistreatment; Track Two is called Alternative Response (AR) and is applied to all self-neglect allegations and lower-risk allegations of mistreatment. The AR practice opens the door for more collaborative engagement by establishing a strong partnership from case start to case end, as illustrated in the pilot theory of change (Figure 1).

As an innovative practice in Colorado and nationally, SB21-118 requires a two-year outcomes evaluation to assess effectiveness of the AR Pilot and inform the future of the dual-track model. The Colorado Department of Human Services (CDHS), Adult Protective Services (APS), partnered with the Colorado Evaluation and Action Lab (Colorado Lab) to fulfill this legislative opportunity. **This report summarizes year one data on pilot reach and implementation and provides preliminary causal evidence on effectiveness of AR at the person and system levels.**

Participating Counties

SB21-118 authorized 15 counties to participate in the AR Pilot, with a requirement for a balance of rural/frontier and urban/suburban counties. Counties applied as interested in participating. The Colorado Lab used a method called random stratified sampling with weighting to select the 15 counties in a fair and balanced manner (Table 1).

Table 1. Participating Pilot Counties and Geographic Designation

Rural/Frontier Counties	Urban/Suburban Counties
Eagle	Adams
Garfield	Arapahoe
La Plata	Denver
Otero	El Paso
Prowers	Jefferson
Pitkin	Larimer
Routt	Mesa
---	Weld

Figure 1. Alternative Response Practice Theory of Change



Description of the Study

The evidence-building approach prioritizes data-informed learning alongside rigorous evaluation methods.

The sample period is January 4, 2023 through December 31, 2023 (implementation year one), with follow-up through June 30, 2024.

Fidelity was measured to ensure pilot counties were delivering the practice as intended to drive outcomes.

Descriptive analysis illustrates reach and implementation of the pilot.

A quasi-experimental design (QED) was used to generate initial causal evidence of effectiveness.

Evidence-Building Approach

Colorado is committed to data-informed state investments and strategic decision making. As a pilot program, it is imperative that research evidence is generated on AR to inform practice and policy development, scalability, and sustainability. To meet this goal, our evidence-building approach maximizes actionability with rigor. During the pilot, the focus is on data-informed learning and strengthening implementation. At the conclusion of a two-year implementation period, the focus is on generating initial causal evidence of effectiveness. In this report, we present select reach, implementation, and impact evidence from the first year of pilot implementation (January 4, 2023 through December 31, 2023), with follow-up through June 30, 2024.

Fidelity of Implementation

Fidelity monitoring is an essential component of the AR Pilot evaluation. Fidelity monitoring helps to answer the question, “Is the pilot being implemented as intended?” Fidelity monitoring explores what activities actually occurred and contributed to outcomes. Fidelity monitoring is essential to continuous improvement and to creating a cohesive, replicable version of the AR practice. Fidelity measurement is a collaborative process between the Colorado Lab, the AR Pilot Planning Specialist, and the CDHS Administrative Review Division. See [Appendix A](#) for details.

Descriptive Analysis

Descriptive analysis is used to understand reach and implementation of the AR Pilot. These analyses are based on cases that were screened in, investigated, and closed between January 4, 2023 and December 31, 2023. In total, there were 4,350 cases, representing 6,511 allegations and 4,069 unique clients. To assess geographic variation, rural-urban county comparisons were done. To explore differences within the dual-track model, comparisons by track assignment (AR versus TR) were done. Tests of statistical significance were conducted throughout.

Quasi-Experimental Design

The outcomes evaluation employs a quasi-experimental design (QED) approach; specifically, a matching approach using propensity scores called inverse probability weighting. Inverse probability weighting is a well-established and vetted procedure in the causal inference toolbox.¹ Grounded in pilot design and the theory of change, this method identifies cases in the pre-pilot period with similar features to AR cases in the pilot period. This information is combined into a single propensity score representing the probability that a pre-pilot case would have an allegation tracked to AR had the dual-track model existed. Cases with similar scores are considered comparable and are weighted more heavily in the analysis. This method reduces subjectivity and improves the precision of causal estimates. More information on the QED methods and variables used for matching are provided in [Appendix B](#).

Defining the Sample

To understand the impact of the AR practice, outcomes of AR cases during the pilot (pilot period: January 4, 2023 through December 31, 2023; tracking outcomes through June 30, 2024) are compared with equivalent cases before the policy lever enabled a dual-track model (pre-pilot period: January 4, 2021 to December 31, 2021; tracking outcomes through June 30, 2022).

The analytic sample is defined as:

- **AR cases:** Pilot period cases that contain one or more AR-tracked allegations (n=2,345).
- **AR equivalent cases (“equivalent cases”):** Pre-pilot period cases that would have at least one AR-tracked allegation had a dual-track model existed and are the strongest matches to AR cases in the pilot period (pool of n=4,302).

Why define AR cases as “one or more allegations?”

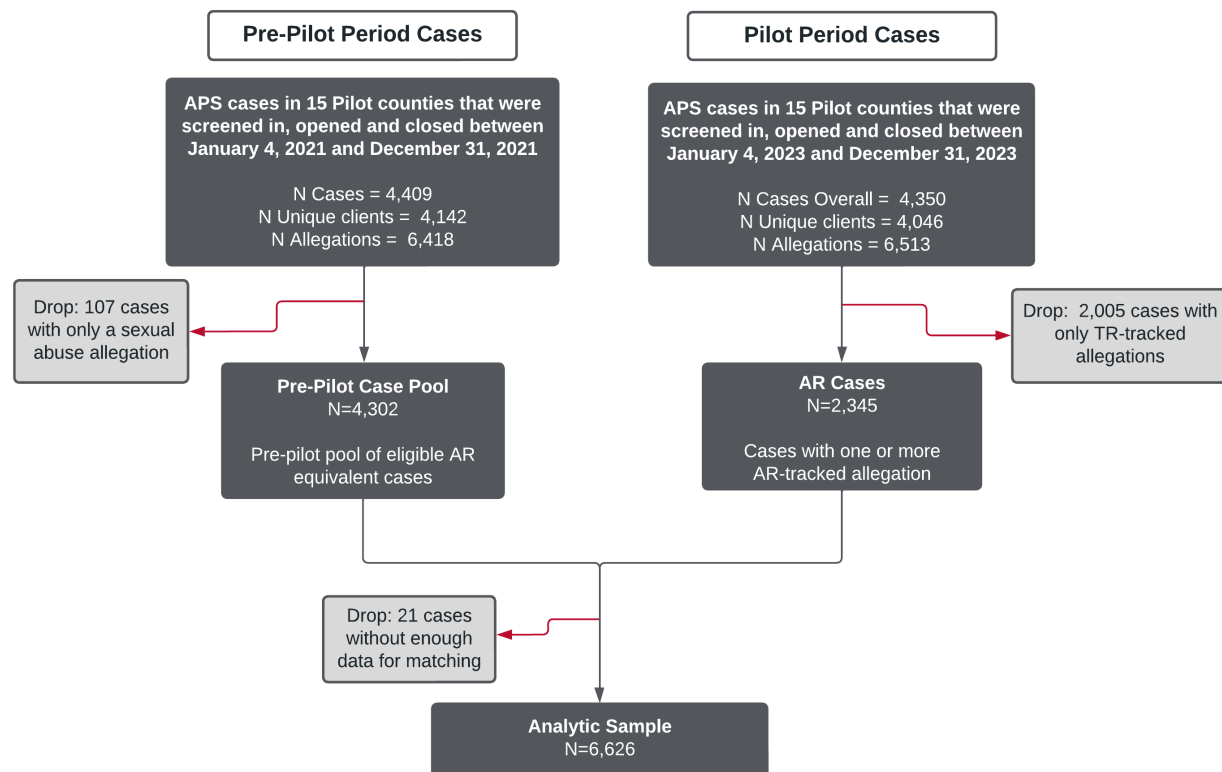
The decision to define “AR cases” as one or more AR allegations (versus cases with only AR allegations) reflects a more rigorous and inclusive approach to assessing impact of the AR practice. Including all cases with at least one AR allegation is the broader and more conservative approach. This allows us to detect whether even one AR-tracked allegation in a case makes a difference.

This approach also allows for more detailed analysis to inform implementation decisions and drive precision practice. For example, how do outcomes differ by cases with only AR allegations, versus those with mixed allegations? By defining the pilot sample as cases with at least one AR allegation, the dual-track model can be understood more holistically. Practically speaking, the majority of AR cases (66.10%) have a single allegation.

Construction of the Analytic Sample

Figure 2 summarizes how pre-pilot and pilot cases are used to construct the analytic sample. The analytic sample is taken from all allegations that were screened in, opened, and closed in the pre-pilot or pilot period. For each of the two periods, a six-month window is added at the end to measure repeat involvement of clients. Cases with only TR-tracked allegations during the pilot are excluded from the analytic sample. An in-depth investigation of pilot period cases with only TR-tracked allegations found that these cases should not be included in the analytic sample. We also excluded 107 cases with only a sexual abuse allegation from the pre-pilot sample because sexual abuse allegations can never be tracked to AR. Finally, 21 cases are lost during the estimation of propensity scores because they do not have valid values for all matching variables. The final analytic sample consists of 6,626 cases.

Figure 2. Construction of the Analytic Sample



Using Qualitative Data for Context

1. Qualitative data from focus groups and feedback from pilot counties help to provide context to results from fidelity, descriptive, and QED analyses.
2. Qualitative data also speak to caseworker and client satisfaction with the AR practice—a valuable outcome in its own right.
3. Narrative findings are integrated throughout.

Key Findings

Fidelity of Implementation

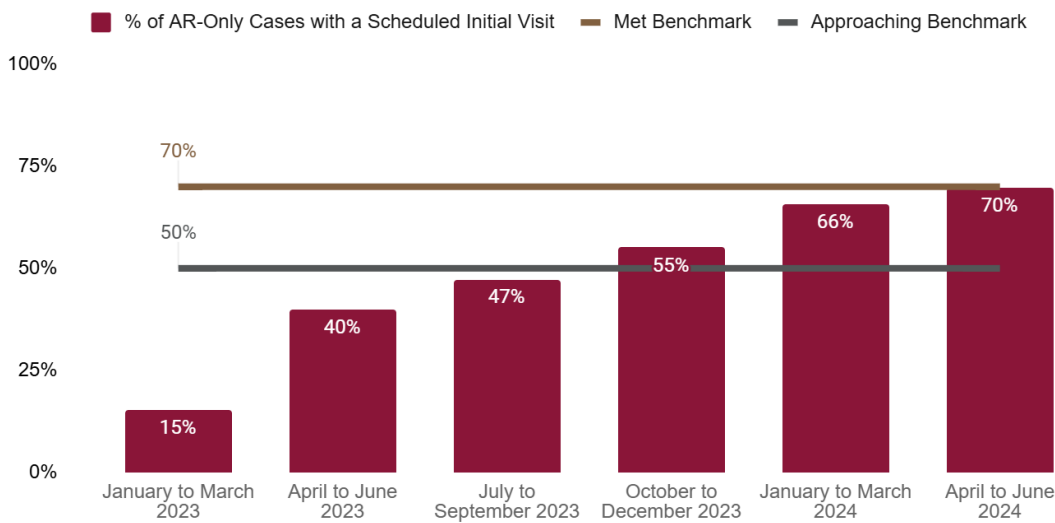
The AR Pilot has seven fidelity measures. Fidelity was measured for the first 12 months of implementation. Measuring fidelity early in the pilot provides opportunity to strengthen implementation. Results show that, on the whole, fidelity was strong in the first year of this novel practice: all pilot counties were approaching or met fidelity in six of the seven measures.¹ [Appendix A](#) details fidelity results for all measures.

Spotlight on Initial Response

Early data from one measure (initial response) indicated that the option to schedule the initial visit was being exercised far less than anticipated and desired to drive outcomes (only 37% of AR-only cases used this option). As such, this metric was selected for in-depth learning and action, and was re-measured for changes over an 18-month pilot period.

The option to schedule the initial visit has steadily improved over time (Figure 3). However, there is still room for improvement. Qualitative data illustrate that moving from an unannounced to a scheduled initial response is a large culture change within APS and requires caseworkers to adopt new skills to effectively implement. This growth curve is reflected in the data, where pilot counties were slow to start in using this hallmark feature of the AR practice. CDHS has provided significant guidance and implementation support to counties toward improving this rate over time. This is critical, as qualitative data show caseworkers appreciate the option to schedule an initial visit and feel that, when exercised, it can help accelerate rapport building.

Figure 3. Scheduled Initial Visits (18-month pilot period: January 2023 through June 2024)



¹ One rural county received a “not met” score on a singular additional measure (use of data), which largely reflects staff availability constraints in attending pilot county meeting.

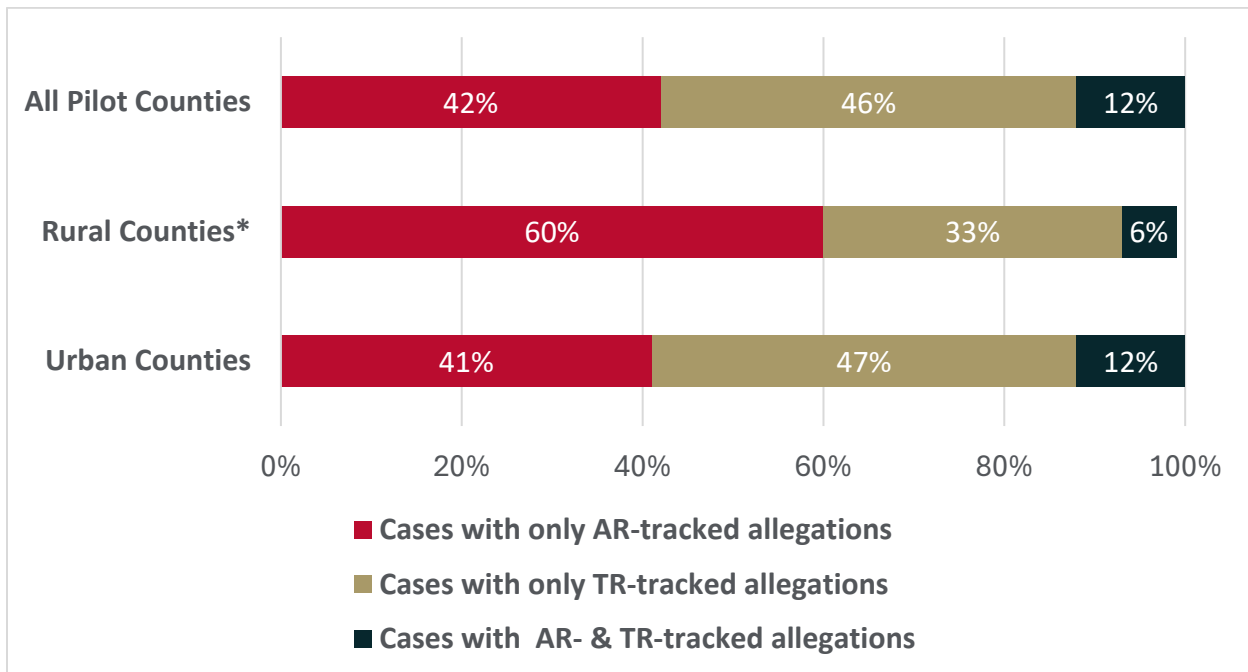
Descriptive: Reach and Implementation

A total of seven indicators were descriptively analyzed, covering reach and implementation of the AR practice in pilot counties.

Insight 1: Just under half of APS cases have only AR-tracked allegations. This is higher in rural counties.

The AR track is being robustly used by pilot counties (Figure 4), signaling the need for a dual-track model that can tailor response to level of risk. Use of the AR track is significantly higher ($p < 0.01$) in rural counties. Qualitative narratives indicate that especially in small-knit communities, the AR track can improve collaboration, particularly for older adults who are strongly independent and desire to age in place. The average age of clients served by pilot counties is 67 years old.

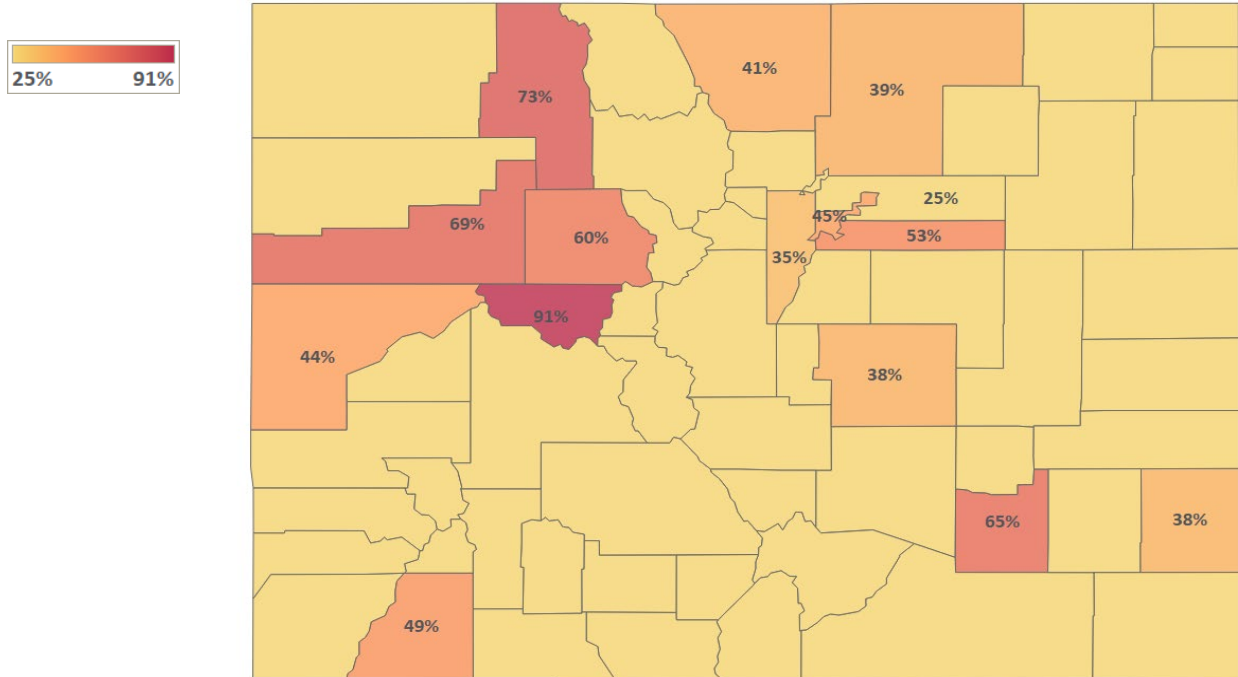
Figure 4. Allegation Breakdown on Cases



* $p < 0.01$ (rural counties compared to urban counties)

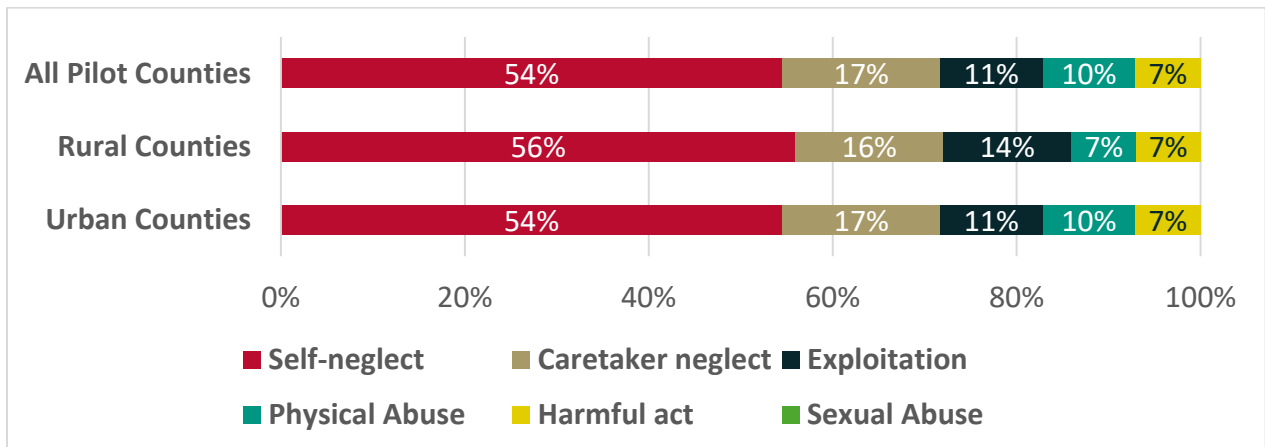
Insight 2: AR Pilot counties show variation in their use of the AR track.

It is important to explore how the dual-track system is unfolding in different contexts, such as rural versus urban (Figure 5). Variation may reflect caseworker discretion, different case complexities, different staffing structures (generalist versus specialist), different service availability, and different caseloads. Importantly, higher use of the AR track is significantly correlated with higher self-neglect allegations (correlation coefficient: 0.7811). This is an anticipated result since, by rule, all allegations of self-neglect are automatically assigned to the AR track.

Figure 5. Use of the AR Track, by county


Insight 3: Self-neglect makes up over half of all AR-tracked allegations. This is slightly higher in rural counties.

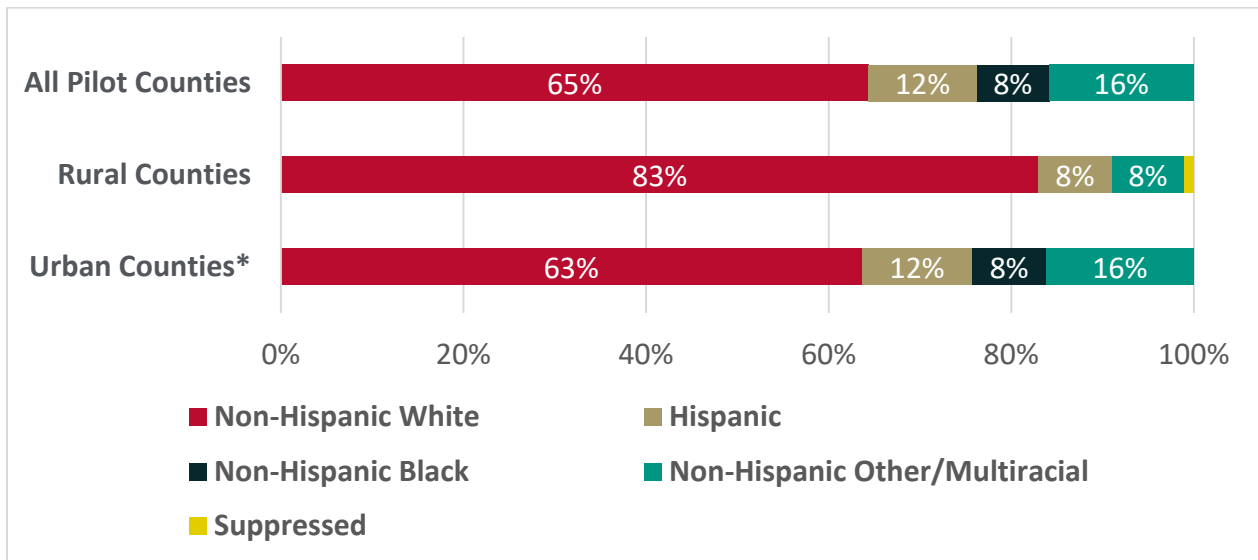
Understanding what is driving track assignment can help inform understanding of the model. Data show that self-neglect makes up over half of all AR-tracked allegations (Figure 6). This aligns with the theory of change and the underlying philosophy of the dual-track system to match response approach to risk level.

Figure 6. Breakdown of Allegations in the AR Pilot


Insight 4: Understanding equitable reach in the AR Pilot

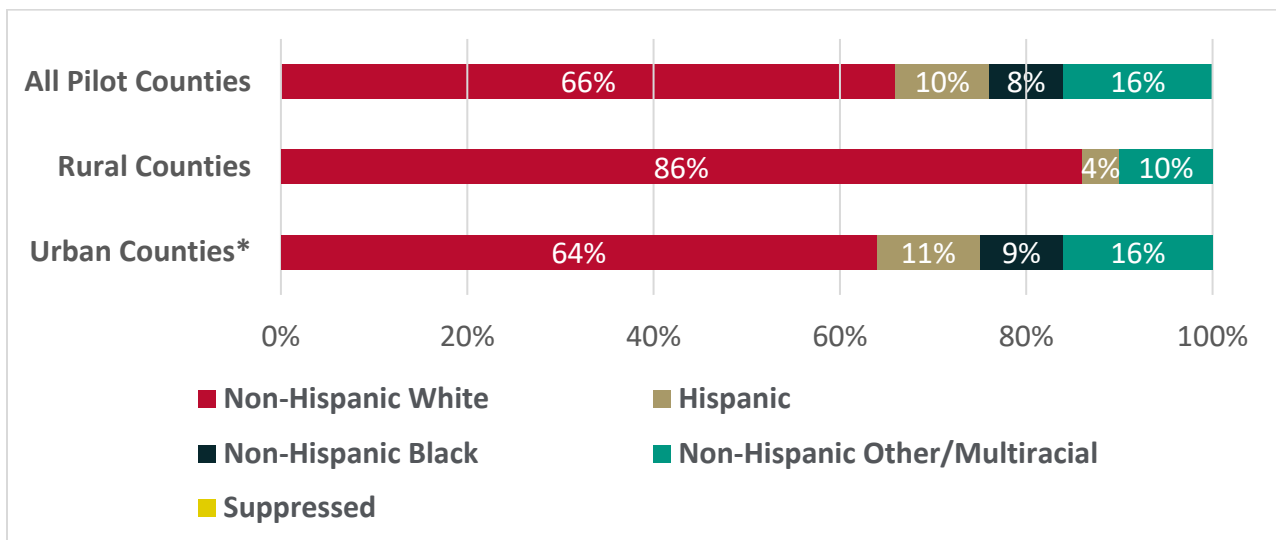
As part of the CDHS commitment to equity, diversity, and inclusion, it is important to understand whether the AR Pilot and the dual-track model are equitably reaching clients. Demographic data in the Colorado Adult Protective Services (CAPS) system is a known limitation, as about 50% of clients have missing race/ethnicity demographic data. For clients with available race/ethnicity data, no disparities are observed in who the AR practice is reaching (Figures 7a and 7b). Urban counties are significantly ($p < 0.01$) more likely than rural counties to serve clients identifying as Black, Hispanic, or multiracial.

Figure 7a. Race and Ethnicity of Clients served by APS in the Pilot Counties (all allegations)



* $p < 0.01$ (urban counties compared to rural counties)

Figure 7b. Race and Ethnicity of Clients served by APS in the Pilot Counties (AR allegations)

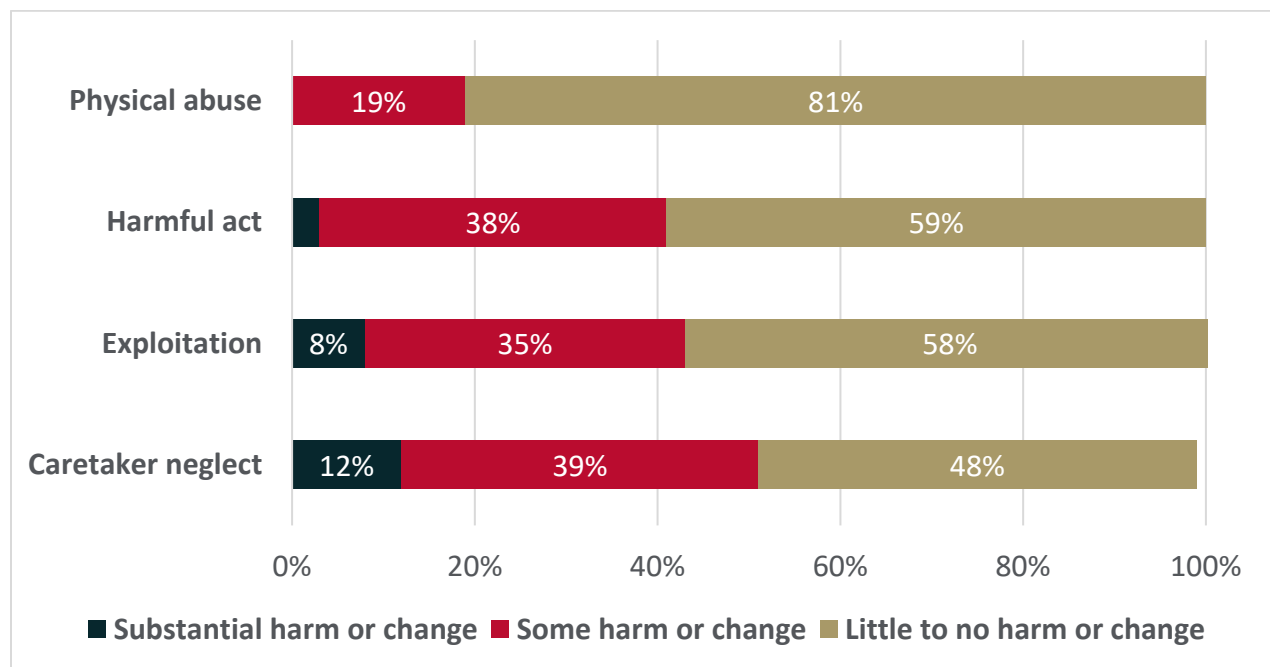


* $p < 0.01$ (rural counties compared to urban counties)

Insight 5: Conclusions for AR-tracked allegations reflect the low-risk nature of the AR track and signal that the AR track is being used appropriately.

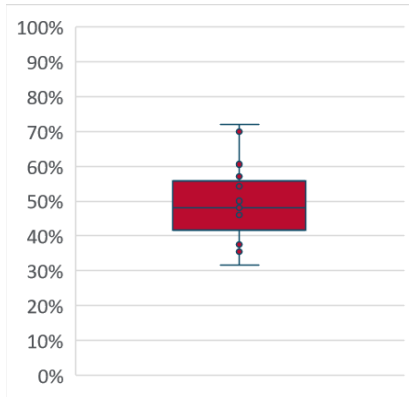
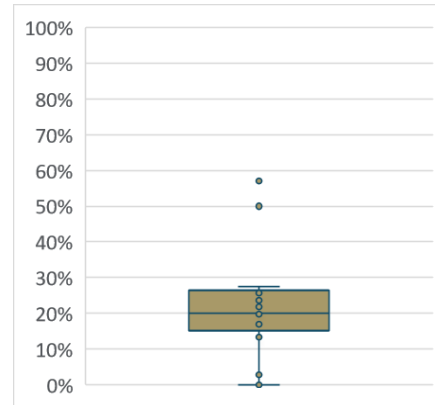
A hallmark feature of AR is that there is no finding for AR allegations, but rather, a conclusion. It is still important to understand the extent of impact that occurs in AR allegations and to assess whether this pattern aligns with the intended low-risk nature of the AR practice. For allegations of mistreatment tracked to AR, conclusion data show an appropriate use of the AR track (Figure 8). All cases of substantial impact have been reviewed by the AR Specialist and confirmed that the track was appropriately used. Qualitative narratives indicated that especially in cases of caretaker neglect involving a spouse, the AR track is well-suited to support not only the client, but the relationship and can stabilize the whole family by addressing root causes of involvement (e.g., husband having difficulty caring for wife with Alzheimer’s. Needs respite twice a week).

Figure 8. Extent of Harm for AR Allegations with a Conclusion



Insight 6: Clients with only-AR tracked allegations are significantly more likely to live alone and have fewer support networks.

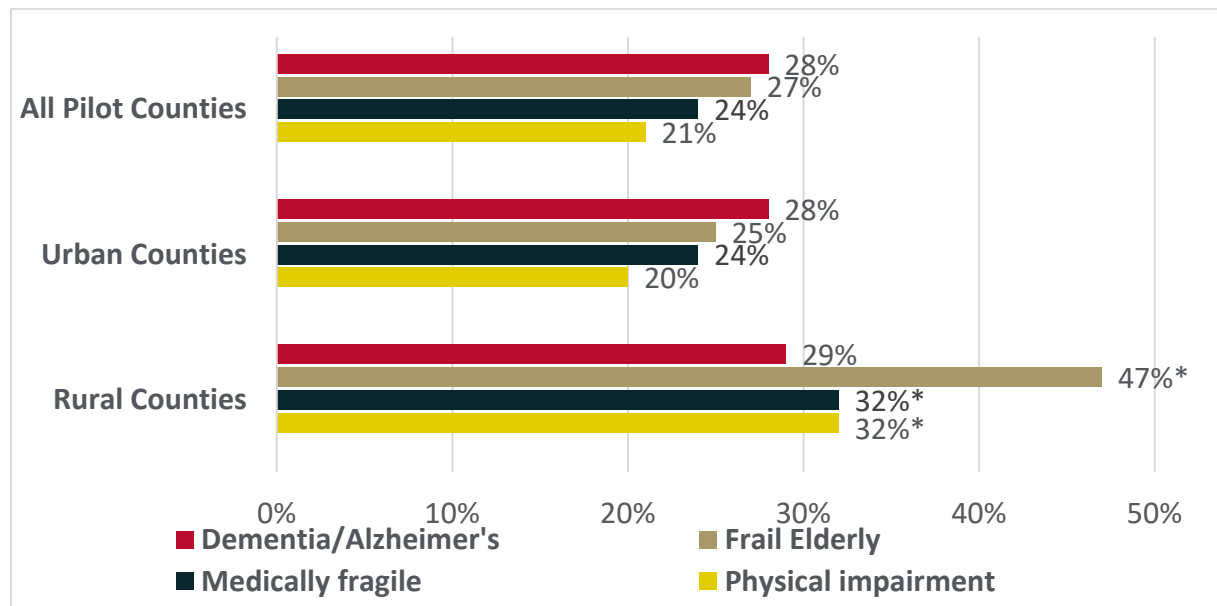
Caseworkers can collaborate with clients to strengthen their support networks and stabilize the client in the home to help prevent future involvement and escalation of mistreatment or self-neglect. Anticipatory practice guidance can be developed from an understanding of client situations at case start. Clients with only AR-tracked allegations are significantly ($p < 0.01$) more likely to live alone (48% county-level median) compared to clients with only TR-tracked allegations (20% county-level median) (Figure 9). Living alone status is a proxy for social isolation. Social isolation is heightened by a lack of support networks, which are significantly ($p < 0.01$) lower for clients with AR-only cases than for clients with TR-only cases (2.61 persons for AR vs. 2.91 persons for TR).

Figure 9. Clients Living Alone, by track type and county
AR-Only Cases

TR-Only Cases


Note: dots represent county-level medians

Insight 7: Leading conditions vary across rural and urban pilot counties.

Understanding client conditions can further inform anticipatory guidance and identify who the AR practice is most appropriate for. Leading conditions also directly interact with social isolation (Insight 6) and can influence both why a client is involved with APS and how best to serve them. In cases with only AR tracked allegations, the leading conditions are dementia/Alzheimer's, frail elderly, medically fragile, and physically impaired (Figure 10). The latter three conditions are significantly higher in rural communities ($p < 0.05$), with frail elderly nearly double compared to urban counties. Qualitative narratives show the AR practice enables caseworkers to better help the client build support networks, which are essential to managing these conditions long-term.

Figure 10. Leading Conditions for AR-Only Cases


* $p < 0.05$ (rural counties compared to urban counties)

Inferential: Confirmatory Outcomes

A total of 13 outcomes were analyzed covering repeat involvement, case length, and client engagement. For repeat involvement, the QED estimates the average difference in the likelihood of a client having a second screened-in case between initial AR cases and initial equivalent cases on three measures. For case length and client engagement, the QED estimates the average difference between AR cases and equivalent cases for six case length and four client engagement measures. Together, these speak to client- and system-level outcomes. Definitions for each measure are included in [Appendix C](#).

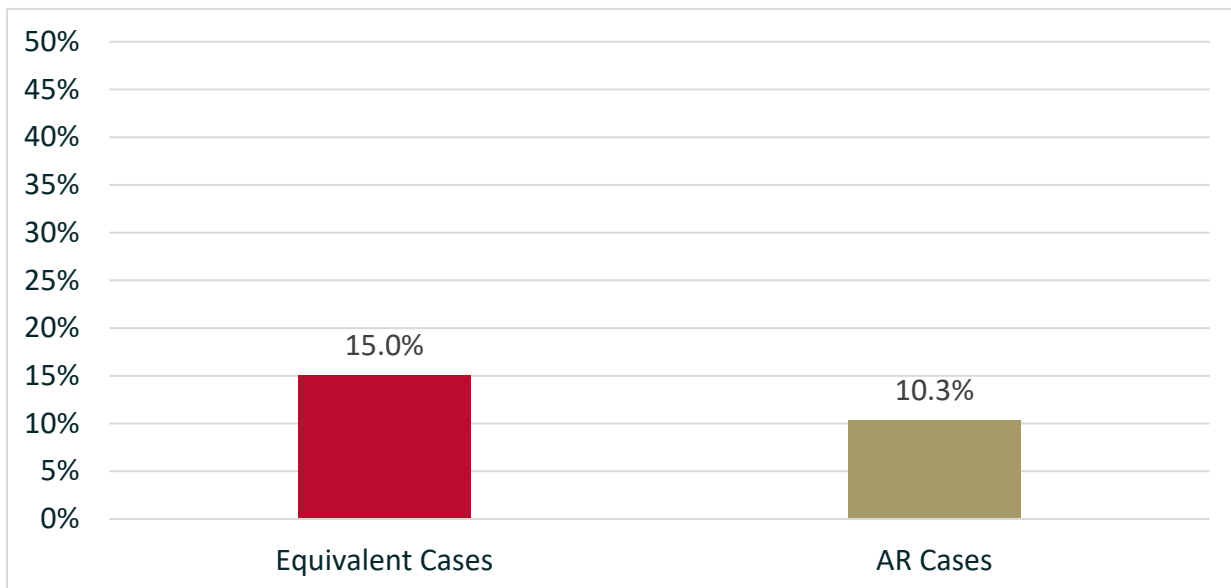
Outcome 1. The Alternative Response Practice Reduces Repeat Involvement

For all three measures of repeat involvement, having at least one AR-tracked allegation on the initial case reduced the likelihood of having a second screened-in case. The estimated differences between AR cases and equivalent cases are all statistically significant. Qualitative narratives indicate that the AR practice can lead to more collaborative case planning, which in turn can help address root causes and reduce the need for repeat involvement for the same issue.

Outcome 1. Repeat Involvement

Finding: Compared to equivalent cases, AR cases were 4.71% less likely to have a second screened-in case ($p < 0.01$). The estimated repeat involvement rate for equivalent cases is 15.01%, and the estimated repeat involvement rate for AR cases is 10.30% (Figure 11).

Figure 11. Comparison of Estimated Rates of Repeat Involvement for Equivalent and AR Cases



Outcome 1a. Repeat Self-Neglect

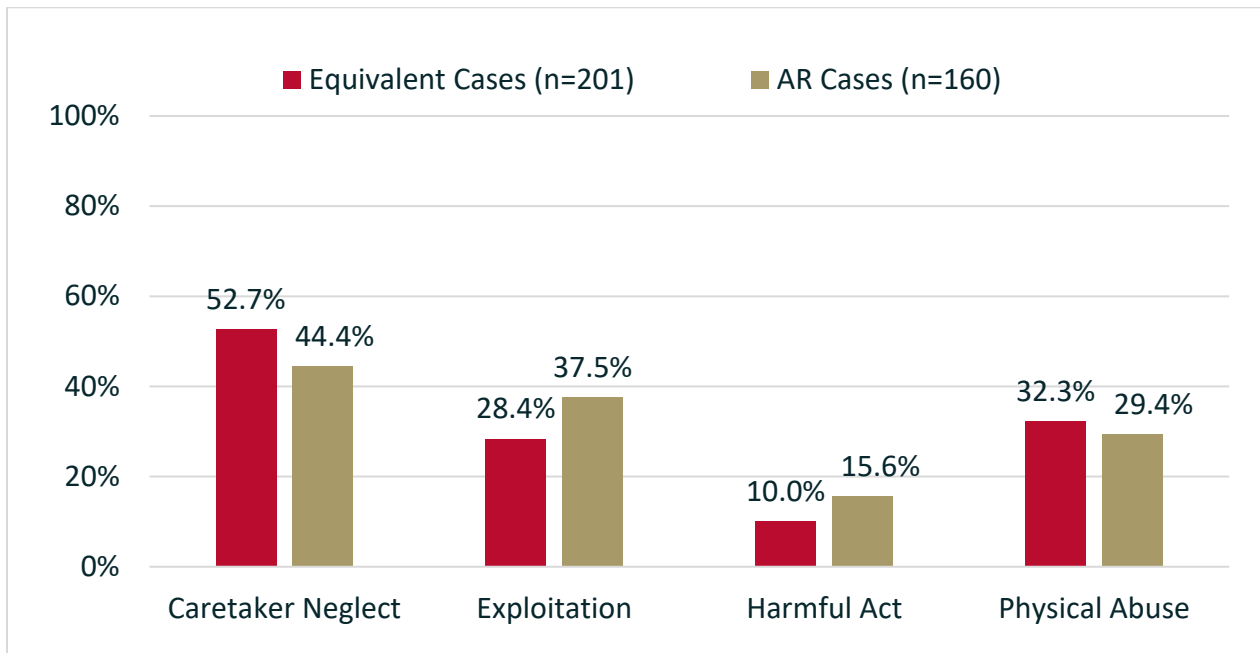
Finding: Of the subset of clients who have a second screened-in case, compared to equivalent cases with a self-neglect allegation, having an AR-tracked self-neglect allegation reduced the probability of having a repeat self-neglect allegation by 10.42% ($p < 0.01$). The estimated probability of repeat self-neglect in equivalent cases is 25.28%, and the estimated probability of repeat self-neglect in AR cases is 14.85%.

Outcome 1b. Repeat Mistreatment

Finding: Of the subset of clients who have a second screened-in case, compared to equivalent cases with a mistreatment allegation, having an AR-tracked mistreatment allegation reduced the probability of repeat mistreatment by 20.05% ($p < 0.01$). The estimated probability of repeat mistreatment in equivalent cases is 33.59%, and the estimated probability of repeat mistreatment in AR cases is 13.54%.

Figure 12 shows the distribution of mistreatment allegations on initial cases from the subset of clients that have a second screened-in case. Equivalent cases have higher percentages of initial cases with an allegation of caretaker neglect and physical abuse, while AR cases have a higher percentage of initial cases with an allegation of exploitation and harmful neglect.

Figure 12. Distribution of Mistreatment Allegations during Initial Involvement for Equivalent Cases and AR Cases with Repeat Involvement

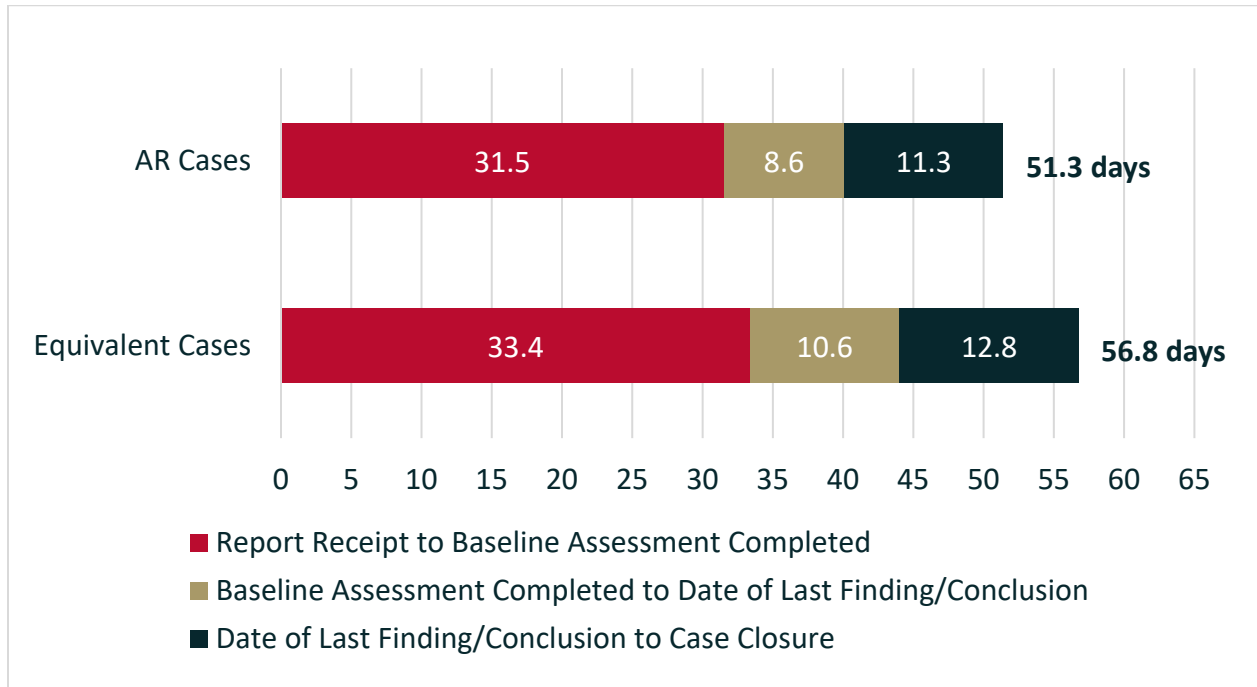


Notes: Cases can have more than one mistreatment allegation, so percentages add to more than 100.

Outcome 2. The Alternative Response Practice Reduces Case Length

Across measures, having at least one AR-tracked allegation shortened the time it took to manage and progress through a case. For AR cases, total case length is reduced by 5.50 days. This reduction comes from quicker completion of client baseline assessments, fewer days in determining findings/conclusions, and a shorter window between the date of last finding/conclusion and case closure. Figure 13 summarizes these changes by showing at which points cases are getting shorter. Qualitative narratives indicate several factors influencing the shorter case timelines. These include improved rapport building using the AR practice that increases trust and can help uncover strengths and needs quicker, as well as a reduced documentation burden that is alleviated through the no finding requirement.

Figure 13. Comparison of Number of Days from Report Receipt to Case Closure



Outcome 2. Total Case Length

Finding: AR cases close 5.50 days earlier compared to equivalent cases ($p < 0.01$). On average, equivalent cases close 56.81 after report receipt days, and AR cases close 51.30 days after report receipt.

Outcome 2a. Report Received Date to Initial Response Date

Finding: For AR cases, initial response occurs 0.04 days later ($p = 0.637$) compared to equivalent cases. On average, initial response occurs 3.46 days after report receipt for equivalent cases and 3.50 days after report receipt for AR cases.

Outcome 2b. Report Received Date to Date Baseline Assessment Completed

Finding: For AR cases, the baseline assessment was completed 1.89 days earlier ($p < 0.01$) compared to equivalent cases. On average, baseline assessments are completed 33.39 days after report receipt for equivalent cases and 31.50 days after report receipt for AR cases.

Outcome 2c. Initial Response Date to Date Baseline Assessment Completed

Finding: For AR cases, the time between initial response to when baseline assessment was completed was 1.87 days earlier ($p < 0.01$) compared to equivalent cases. On average, the time between initial response and baseline assessment completion was 29.89 days for equivalent cases and 28.01 days for AR cases.

Outcome 2d. Report Received Date to Date of Last Finding/Conclusion

Finding: For AR cases, the date of last finding/conclusion on a case was determined 3.95 days earlier ($p < 0.01$) compared to equivalent cases. On average, the last finding/conclusion occurred 44.00 days after report receipt for equivalent cases and 40.05 days after report receipt for AR cases.

Outcome 2e. Initial Response Date to Date of Last Finding/Conclusion

Finding: For AR cases, the time between initial response to date of last finding/conclusion on a case was determined was 3.99 days earlier ($p < 0.01$) compared to equivalent cases. On average, the time from initial response to date of last finding/conclusion occurred was 40.54 days for equivalent cases and 36.55 days for AR cases.

Outcome 3. The AR Practice has the Potential to Improve Client Engagement

Across measures, the estimated differences between AR cases and equivalent cases were found to be statistically insignificant and small in magnitude. The only significant result shows that AR cases experienced a smaller reduction in the percent of services refused compared to equivalent cases. Qualitative narratives and the theory of change help shed light on the client engagement outcomes. Caseworkers consistently report that a major value of the AR practice is the ability to engage collaboratively from case start. Further, they report that for self-neglect and low-risk mistreatment allegations, the AR practice is more person-centered. However, as noted in the [fidelity section](#), pilot counties were slow to start on the hallmark feature of option to schedule an initial response. Given county improvements in this implementation practice over time, we anticipate the two-year outcome data may show more significant differences in client engagement outcomes. A full two-year sample will also increase statistical power by including more cases, which will further help improve the ability to detect change.

Outcome 3a. Client Refusing Contact

Finding: The average difference in clients refusing contact between AR cases and equivalent cases is 0.02% ($p=0.978$). On average, the probability of a client refusing contact for equivalent cases is 4.10% and 4.11% for AR cases.

Outcome 3b. Client Refusing All Services

Finding: Compared to equivalent cases, AR cases are 0.83% less likely to have a client refuse all services ($p=0.177$). On average, the probability of a client refusing all services for equivalent cases is 6.02% and 5.19% for AR cases.

Outcome 3c. Percent of Services Refused

Finding: Compared to equivalent cases, clients with AR cases refuse 1.30% less services ($p<0.05$). On average, 8.88% of services are refused in equivalent cases and 7.58% in AR cases.

Outcome 3d. All Services Ineffective

Finding: Compared to equivalent cases, AR cases are 0.05% less likely to have all services deemed ineffective ($p=0.709$). On average, the probability of all services being ineffective for equivalent cases is 0.32% and 0.27% for AR cases.

Inferential: Exploratory Outcomes

Exploratory analysis dives deeper into the role of client support networks on client engagement and track changes on case length.

Client Engagement Can Be Improved through Support Networks

Larger support networks can improve client engagement in key ways. Clients are less likely to refuse contact when their support network is larger. For each additional support, the likelihood of a client refusing contact goes down by 1.20% ($p<0.01$). When examining refusal of services, a larger support network reduces the likelihood of a client refusing all services. For each additional support, the likelihood a client refused all services goes down by 0.37% ($p<0.05$). Qualitative narratives indicate these results may reflect improved trust that comes when support networks are included in the APS response and collaborative case planning with clients.

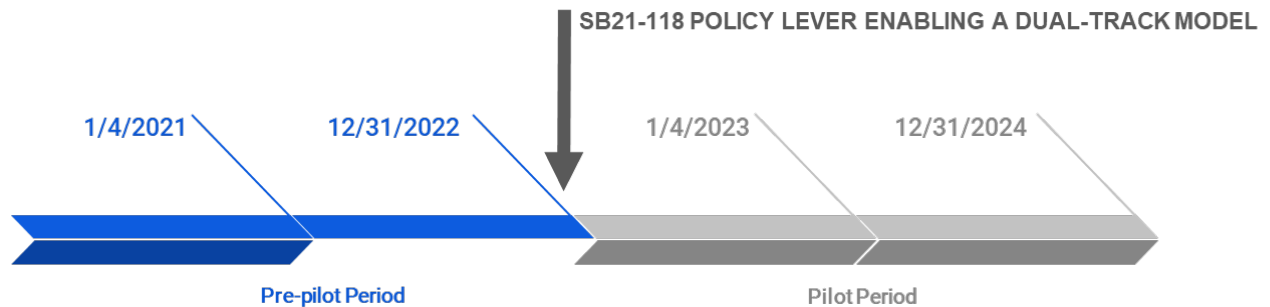
Tracked Changes Increase Case Length

Track changes are very uncommon, with only 87 in the sample (3.71% of all AR cases). Across [all six measures of case length](#), having a track change increased the days it took to progress through a case. The findings were all statistically significant ($p<0.01$), with the exception of time between report receipt to initial contact. Overall, a track change increased the length of an AR case (report receipt to case closure) by 7.35 days ($p<0.01$). This is an expected finding. Qualitative narratives show that caseworkers appreciate the option to change tracks as new details emerge in the case and that a slightly longer case length is a worthwhile trade-off.

Implications

Initial results from the QED indicate the AR Pilot is having a positive impact on at-risk adults in Colorado, specifically by reducing repeat involvement and case length. Causal evidence generated by this rigorous evaluation will be updated (and outcomes measured will be expanded), once the full two years of implementation data become available (Figure 14).

Figure 14. Outcomes Comparison for a Two-Year Pilot Period



**A six-month follow-up period will be added to both time periods.*

Together, qualitative and quantitative data indicate the AR practice is a viable and favorable approach for responding to reports of low-risk mistreatment and self-neglect. Assuming causal evidence remains favorable in the full two-year outcomes study, the dual-track model should be recommended for statewide expansion. A phased rollout will be critical to ensure state and county APS have the time, resources, and training necessary to implement the AR practice to fidelity and achieve outcomes documented through the evaluation.

Finally, while evidence building for AR is focused firstly on APS response, evaluation results also have implications for the aging population across units at CDHS, including the State Unit on Aging. As Colorado and the nation grapple with how best to care for this rapidly growing community, it is imperative that prevention and intervention services are in place that reflect the unique conditions and challenges of this population. For example, in-depth data on self-neglect generated through this pilot can inform best practices for the aging population and clearly identifies that more resources are necessary across prevention levels.

Appendix A. Fidelity Measures

12-Month Fidelity Assessment

The Colorado Lab initially assessed fidelity using data from the first eight months of the pilot (January 2023 through September 2023). Select fidelity measures were reassessed in January 2024. Table A1 shows the results of this fidelity assessment. “N/A” applies if there were no cases to review that met the criteria for the indicator at the time of measurement.

Table A1. Fidelity of Implementation Ratings by County and Indicator, January 2024

County	1. Initial Track Assignment	2. Initial Response	3. Track Changes	4. Investigation and Conclusion	5a. Matching Needs to Services AR only cases	5b. Matching Needs to Services Cases with AR and TR allegations	6. Use of Data	7. Continuing Education / Professional Development
Adams	Met	Met	Met	Met	Met	Met	Met	Met
Arapahoe	Met	Approaching	Met	Met	Met	N/A	Met	Met
Denver	Met	Approaching	Met	Met	N/A	N/A	Met	Met
Eagle	Met	Met	Met	Met	N/A	N/A	Met	Met
El Paso	Met	Approaching	Approaching	Met	Met	Met	Met	Met
Garfield	Met	Not Met	Approaching	Met	Met	Approaching	Met	Met
Jefferson	Met	Not Met	Met	Met	Met	Met	Met	Met
La Plata	Met	Not Met	Met	Met	N/A	Met	Approaching	Met
Larimer	Met	Not Met	Approaching	Met	Met	Met	Approaching	Met
Mesa	Met	Approaching	Met	Met	Met	Met	Met	Met
Otero	Met	Approaching	Met	Met	Met	N/A	Approaching	Met
Pitkin	Met	Met	Met	Met	N/A	Met	Met	Met
Prowers	Met	Met	Met	Met	Met	Met	Approaching	Met
Routt	Met	Not Met	Met	N/A	Met	N/A	Not Met	Met
Weld	Met	Approaching	Met	Met	N/A	N/A	Met	Met

Fidelity Measures

1. **Initial Track Assignment:** Evidence that initial track assignment is consistently and appropriately applied by leads and supervisors. Met: >90%; Approaching: 70% to 90%; Not Met: <70%.
2. **Initial Response:** Evidence that the option to schedule an initial visit is being exercised consistently and appropriately. Met: 70%; Approaching: 50% to 70%; Not Met: <50%.
3. **Track Changes:** Evidence that use of the track change option is being exercised judiciously and consistently. Met: >70%; Approaching: 50% to 70%; Not Met: <50%.
4. **Investigation and Conclusion:** Evidence that a determination of a conclusion is being consistently and robustly applied. Met: >70%; Approaching: 50% to 70%; Not Met: <50%.
5. **Matching Needs to Services:** Evidence that services in the case plan are being matched to client needs and their families. Met: >70%; Approaching: 50% to 70%; Not Met: <50%.
6. **Use of Data:** Evidence that data are being used to improve APS practice, drive outcomes, and assure the equitable reach of AR. Met: >80%; Approaching: 60% to 80%; Not Met: <60%.
7. **Continuing Education/Professional Development:** Evidence that county staff participate in mandatory and voluntary trainings and professional development opportunities on AR and apply training to practice and CAPS use. Met: >80%; Approaching: 60% to 80%; Not Met: <60%.

Appendix B: Quasi-Experimental Design Description and Matching Variables

Quasi-Experimental Design (QED) Description: The QED is a matching design with propensity scores called inverse probability weighting. Inverse probability weighting (IPW) is a well-established and vetted procedure in the causal inference toolbox.² The IPW design compares Alternative Response (AR) cases in the pilot period with equivalent cases in the pre-pilot period by identifying which pre-pilot cases have similar features as pilot AR cases. To determine which pre-pilot cases are considered valid comparisons, the approach uses a statistical technique that first estimates the probability a case contains an AR-tracked allegation using a set of matching variables, and then assigns that case a weight (or “importance”) based on its strength as a comparison. Propensity scores is a dominant matching paradigm and tested method for aggregating information from multiple matching variables into a single value.^{3, 4} IPW has the benefit of producing the most accurate causal estimates possible given the constraints posed by data availability and the pilot implementation, while minimizing subjectivity in design decisions.

Through the calculation of propensity scores and their weighting, the QED identifies pre-pilot cases that would have an allegation tracked to AR had the dual-track model been available. These cases are labeled as “equivalent cases.” Cases with AR-tracked allegations from the pilot period are compared to equivalent cases, resulting in an estimated average difference between AR cases and equivalent cases. Propensity scores are used to validate baseline equivalence of cases with AR-tracked allegations in the pilot period and equivalent cases from the pre-pilot period. Balance tests ensure that AR cases and equivalent cases are similar on average.

Selection of Matching Variables: When specifying the logit model to estimate the propensity scores at the case level, we aimed to include variables that determine whether a case has at least one allegation on the AR track. The variables used need to consider the available data and the baseline inputs from the logic model, including a variety of client characteristics and client conditions. The matching variables are predictive variables that determine whether a case had an AR-tracked allegation and are measured at intake and, therefore, not impacted by the intervention itself.^{5, 6} Apart from race/ethnicity, we do not consider variables with mostly missing values (e.g., client income source and health insurance type) or “live” variables that could be updated after intake (e.g., specific at-risk conditions). To augment the case and client variables, we create variables to capture dynamic interactions between age and other factors.

The more observable variables we incorporate, the more likely we account for unobservable variables, thus increasing our ability to assert that we are estimating a causal effect without bias. That said, the use of too many variables increases the likelihood of including colliders or irrelevant variables. Importantly, variables included as matching variables cannot be subsequently used to examine the impact of the dual-track model because the IPW process generates a sample that will be roughly similar along those matching variables. In other words, the differences between AR cases and equivalent cases in these variables will be mechanically as close to zero as possible. Table B1 provides a description of matching variables.

Table B1. Variables Used in the Matching Process for the Quasi-Experimental Design

Category	Variable Description	Notes
Case Characteristics		
	Total number of allegations on a case	
	Types of mistreatment allegations in the case	Indicators for the presence of each of the five mistreatment types
	Presence of a self-neglect allegation on a case	
Client Characteristics		
	County of reporting	
	Age	
	Gender	
	Race/Ethnicity	Missing values coded and labeled as “missing”
	Primary language	
	Presence of a physical, mental, or behavioral condition	
Constructed Variables		
	Quadratic age term	Captures non-linear impact of age x age
	Interaction terms between age and each mistreatment type	Five separate interaction terms
	Interaction term between age and presence of a self-neglect allegation	
	Interaction terms between age and presence of a condition	

Appendix C: Outcome Measures

The 13 client- and system-level outcomes are defined below.

- *Repeat involvement*: A repeat involvement occurs when a client has an initial screened-in case opened and closed within the calendar year (January 4 through December 31) and has a second screened-in case opened within six months of the first case closing.
- *Repeat self-neglect*: A repeat involvement case in which both the first and second cases had an allegation of self-neglect.
- *Repeat mistreatment*: A repeat involvement case in which both the first and second cases had an allegation of mistreatment.
- *Total case length*: The number of days between when a report was received and when the case was closed.
- *Report received to initial response*: The number of days between when a report was received and when a caseworker either made initial contact or attempted to make initial contact, in person or over the phone.
- *Report received to baseline assessment completed*: The number of days between when a report was received and when the baseline assessment was completed.
- *Initial response to baseline assessment completed*: The number of days between a caseworker's initial response (including attempts) and when the baseline assessment was completed.
- *Report received to last finding/conclusion*: The number of days between when a report was received and when a caseworker determined the last finding or conclusion on a case. This is regardless of when allegations were added (i.e., analysis is not restricted to initial allegations only).
- *Initial response to last finding/conclusion*: The number of days between a caseworker's initial response (including attempts) and when a caseworker determined the last finding or conclusion on a case. This is regardless of when allegations were added (i.e., analysis is not restricted to initial allegations only).
- *Client refused contact*: The case was closed because the client refused further contact.
- *Client refused all services*: A {Yes, No} binary indicator for if the client refused all services offered, based on a client's case closure reason.
- *Percent of services refused*: The number of services a client refused out of the total number of services offered in the case plan. A client not offered services received a count of 0.
- *All services ineffective*: A {Yes, No} binary indicator for if all the services provided were determined ineffective, based on a client's case closure reason.

Endnotes

- ¹ Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5-86.
<https://doi.org/10.1257/jel.47.1.5>
- ² Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5-86.
<https://doi.org/10.1257/jel.47.1.5>
- ³ Huntington-Klein, N. (2022). *The effect: An introduction to research design and causality* (1st ed.). Chapman and Hall/CRC Press.
- ⁴ Cunningham, S. (2021). *Causal inference: The mixtape* (1st ed.). Yale University Press.
- ⁵ Heckman, J., & Navarro-Lozano, S. (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *The Review of Economics and Statistics*, 86(1), (30-57). <https://doi.org/10.1162/003465304323023660>
- ⁶ Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5-86.
<https://doi.org/10.1257/jel.47.1.5>