Albemarle County Department of Social Services
APSART Validity Study: Results

Executive Summary

The purpose of this study was to validate the Adult Protective Services Assessment of Risk Tool (APSART), developed by Albemarle County Department of Social Services (ACDSS). The APSART was designed to provide a quantitative risk score, which translates to a risk level (defined as None, Low, Moderate, High), for adults about whom ACDSS has received a report of alleged maltreatment. The overall risk score is the combined value of eight sub-category scores: physical health, mental health, cognitive functioning, environmental risks, support system, abuse/neglect/exploitation factors, economic resources, and alleged perpetrator. Adult Protective Services (APS) Case Workers use the APSART score to assist with decision-making about how to respond to a given report, as well as to monitor improvement in client welfare.

Objectives

We decided to test the APSART's validity based on its ability to predict a repeat APS report following a previous APS report. We analyzed repeat reports within 3, 6, and 12 months of the original report. We set out to answer the following research questions:

1. Does the proportion of clients with repeat APS reports within each APSART risk category increase as the risk level increases? Does this vary by type of allegation?

2. Is the APSART score associated with the likelihood of a repeat APS report?

3. Does the APSART's association with repeat reports vary by ACDSS caseworker?

Methods

Study Population

ACDSS began administering APSART to clients for whom an APS report had been filed in 2010. We started with data on all clients for whom an APS report was filed during December 1, 2010-May 31, 2013 (N=811 individuals). Some clients had as many as ten APS reports during that time period. For our analysis, we restricted this dataset to only those clients with a report that occurred during the period of December 1, 2010-February 28, 2013, to ensure that all clients being included had at least 3 months of follow-up time. To assess repeat reports within 6 and 12 month time windows, we restricted our dataset to only those reports made during the periods of December 1, 2010-November 30, 2012 and December 1, 2010-May 31, 2012, respectively. ACDSS's protocol is to complete the APSART for all clients for whom

an APS report has been filed, although we have not always been consistent in this practice. Thus, we also excluded any clients with a report during that time period who had a missing or "deceased" APSART disposition score. This left us with a total study population of 387 APS clients for the 3-month follow-up period, 375 for the 6-month follow-up period, and 289 for the 12-month follow-up period. For the sake of space, we will primarily report the results of the 6-month follow-up period analyses.

Analysis

APSART data (worker name, score, risk category, disposition) are stored on an Excel spreadsheet. We obtained demographic data (client age, race-ethnicity, sex, education level), data on prior ACDSS services (a single dichotomous variable indicating whether the client had any prior referrals), and alleged and substantiated abuse types from the ASAPS database. These two datasets were linked based on the individual case number to create one analysis dataset.

We analyzed the data using SPSS 17.0 and considered p-values less than 0.05 to be statistically significant. We conducted a chi-square test for trend to determine whether the proportion of clients with a repeat report was statistically different among APSART risk categories. A score of ≥80 is categorized as "high", 30-70 is categorized as "moderate", 10-20 is categorized as "low", and 0 is categorized as "none".

We also used logistic regression to analyze the association between APSART score and at least one repeat report within a given time window (3, 6, or 12 months). We assessed the tool's predictive power by calculating the area under the receiver operator characteristic (ROC) curve for each regression model. By definition, an AUROC=0.5 indicates predictive power that is equal to chance; higher values indicate higher predictive power.

Finally, we examined the APSART's reliability among different workers by creating a series of fictional reports for which workers completed the APSART. Seven APS workers independently completed the APSART for the same five case scenarios. We calculated the intraclass correlation coefficient (ICC) for the workers' overall risk scores, as well as the sub-category scores.[1] We considered an excellent ICC to be 0.75-1.00, good ICC to be 0.60-0.74, fair ICC to be 0.40-0.59, and a poor ICC to be less than 0.40 (Cicchetti, 1994).

Results

APSART Validity Testing

Our study population was primarily non-Hispanic white (80.3%) and female (67.0%), with a mean age of 71 years (Table 1). The mean initial APSART disposition score was 75, which falls between the

---

[1] For all ICC analyses, we calculated a two-way mixed, absolute, average-measures ICC (Hallgren, 2012).

"moderate" and "high" risk levels. Only 11.7% of our study population had an initial APSART risk level in the "none" or "low" risk levels; 46.3% of subjects were categorized as moderate risk and 42.1% were categorized as high risk. This skewed distribution is somewhat expected because an APS report is only filed after all the invalid referrals have been filtered out.

More than one-third of clients with a report during the study period had a history of at least one prior APS report (35.4%). The proportion of clients with a previous APS report increases as the APSART risk level increases, from 0.0% among low-risk reports to 46.6% among high risk reports.

Table 1. Sample characteristics, overall and by initial APSART risk level. Unless otherwise noted, values are n(%).

| Characteristic	Overall	APSART Risk Level			
		None	Low	Moderate	High
Total*	387 (100)	6 (1.6)	39 (10.1)	179 (46.3)	163 (42.1)
APSART Disposition Score - mean (SD)	74.7 (46.5)	0.0 (0.0)	16.4 (4.9)	50.6 (13.9)	117.9 (37.2)
Age - mean (SD)	70.9 (18.2)	76.8 (10.1)	74.8 (16.4)	71.5 (17.4)	69.2 (19.4)
Sex - Female	258 (67.0)	5 (83.3)	26 (68.4)	120 (67.41)	107 (65.6)
Race/Ethnicity - Non-Hispanic White	286 (80.3)	5 (100.0)	25 (71.4)	132 (81.0)	124 (81.0)
Previous APS Report - Yes	137 (35.4)	0 (0.0)	6 (15.4)	55 (30.7)	76 (46.6)

Excludes all entries with a missing or "deceased" disposition score, or report date not during 12/1/2010-2/28/2013.

We found a statistically significant dose-response relationship between the initial APSART risk level and the proportion of clients with a repeat report within 6 and 12 months (Figure 1a). Among initial reports with a low risk score, 12.8% had a repeat APS report within 6 months, compared to 19.8% of reports with a moderate risk score and 24.7% of reports with a high risk score. There was a similar trend for repeat reports within 12 months: 7.1% of low risk reports had a repeat APS report within 12 months, compared to 26.2% among moderate risk reports and 32.3% among high risk reports. Although the proportions changed, the trends remained similar and significant when we analyzed the proportion of initial reports with a *substantiated* repeat report within 6 and 12 months (Figure 1b).

Figure 1a. Among APS reports during 12/1/10-2/28/13, the proportion with a repeat APS report within the given time window, by initial report's APSART risk level.
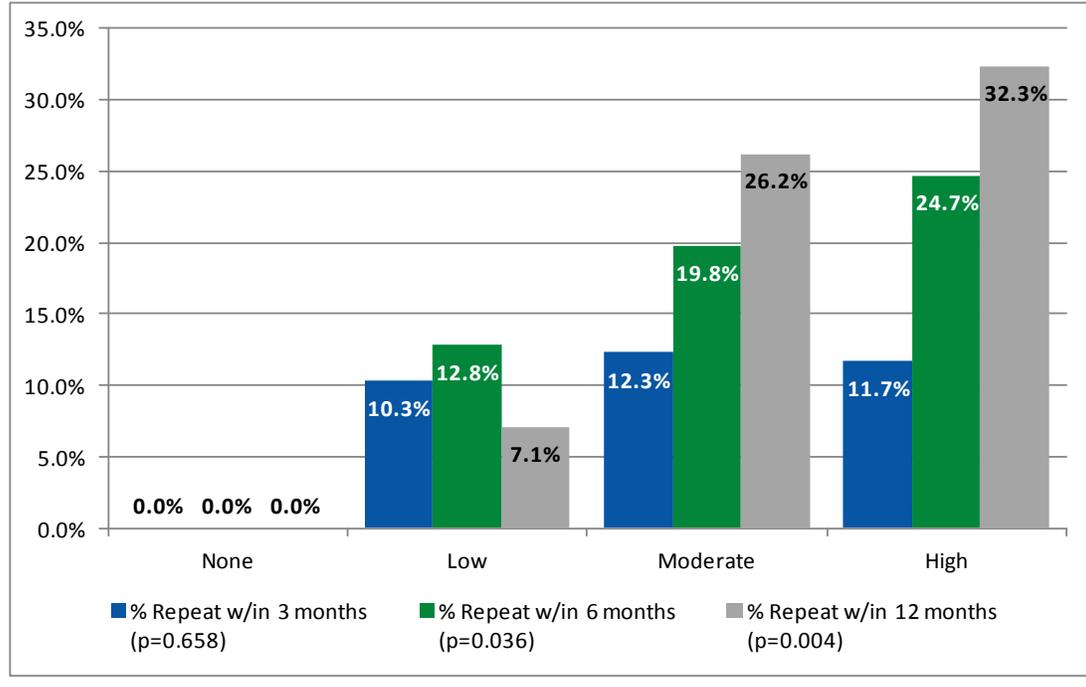
Figure 1b. Among APS reports during 12/1/10-2/28/13, the proportion with a substantiated repeat APS report within the given time window, by initial report's APSART risk level.
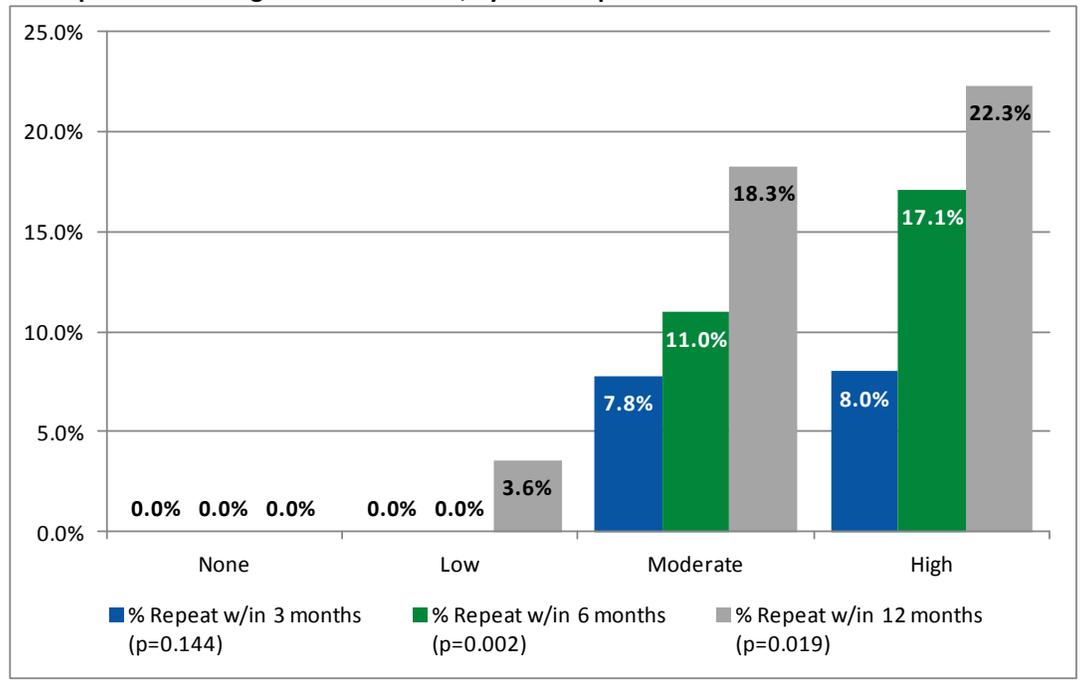
We did not find many statistically significant trends when we assessed the proportion of repeat reports by the characteristics of the initial report (Table 2). This is most likely at least partially due to small sample sizes in each sub-category. Patterns were similar among sub-categories for the 3-month or 12-month follow-up periods.

The logistic regression analysis revealed that the overall APSART risk score is significantly associated with an increased likelihood of a repeat report within 6 or 12 months (Odds Ratio 1.01 each, beta=0.006 and 0.007, respectively). For the 6-month follow-up period, a 10-point increase in APSART risk score was associated with a 6% increase in the odds of a repeat report, given the same worker. For the 12-month follow-up period, there is a 7% increase in the odds of a repeat report, given the same worker, for every 10-point increase in APSART risk score. The area under the receiver operator characteristic (ROC) curve indicates that, when controlling for the worker, the APSART score has fair predictive power regarding the probability of a repeat report within 6 or 12 months (AUROC=0.601 and 0.628, respectively; Figure 2a-b).

Table 2. APS clients with a repeat report of maltreatment within 6 months following a previous APS report, by previous disposition, abuse type, and APSART risk level. Values are n(%).

Previous Disposition/Abuse Type	APSART Disposition Risk Level			p-trend
	Low	Moderate	High	
Overall	5 (12.8)	34 (19.8)	39 (24.7)	**0.036**
Disposition Score				
Unfounded	3 (100.0)	10 (41.7)	9 (29.0)	**0.037**
Need no longer exists	0 (0.0)	11 (45.8)	11 (35.5)	0.864
Need exists - accepts services	0 (0.0)	1 (4.2)	7 (22.6)	**0.046**
Need exists - refuses services	0 (0.0)	2 (8.3)	4 (12.9)	0.427
Type of Allegation*				
Abuse (Mental, physical, and/or sexual)	1 (20.0)	7 (20.6)	6 (15.4)	0.597
Neglect - Other	1 (20.0)	9 (26.5)	18 (46.2)	0.066
Neglect - Self	3 (60.0)	14 (41.2)	19 (48.7)	0.910
Exploitation (Financial and/or other)	0 (0.0)	9 (26.5)	5 (12.8)	0.597
Type of Substantiated Maltreatment*				
Any substantiated maltreatment	2 (40.0)	22 (64.7)	28 (71.8)	0.193
Abuse (Mental, physical, and/or sexual)	1 (20.0)	2 (5.9)	6 (15.4)	0.535
Neglect - Other	0 (0.0)	3 (8.8)	9 (23.1)	0.055
Neglect - Self	1 (20.0)	13 (38.2)	15 (38.5)	0.605
Exploitation (Financial and/or other)	0 (0.0)	6 (17.6)	2 (5.1)	0.367

Subcategories are not mutually exclusive; one individual may have more than one form of alleged or substantiated maltreatment

Figure 2a. ROC curve for the outcome of repeat report within 6 months, as predicted by the initial APSART risk score and worker.
AUROC=0.601 (95% CI: 0.528,0.674); p=0.006.
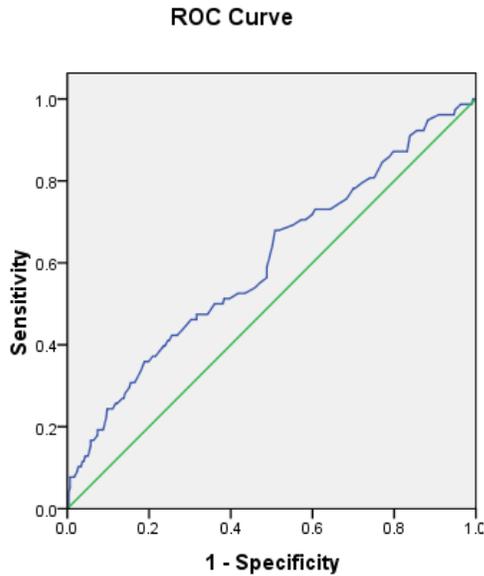
Figure 2b. ROC curve for the outcome of repeat report within 12 months, as predicted by the initial APSART risk score and worker.
AUROC=0.628 (95% CI: 0.555,0.701); p=0.001.
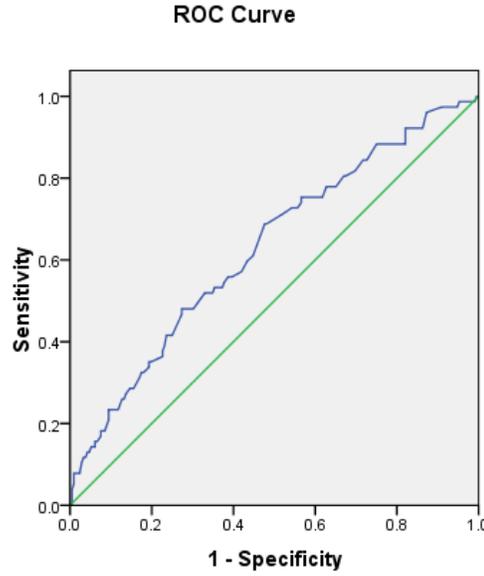
Inter-rater Reliability Testing

There is noticeable variation in the distribution of APSART risk scores assigned by each worker (Table 3). Due to differing skill sets and seniority among workers, this may partly be due to purposeful decisions to assign more difficult cases to specific workers.

Table 3. APSART level at disposition, by worker.

Worker	Overall	None	Low	Moderate	High
1	44 (11.4)	1 (2.3)	2 (4.5)	13 (29.5)	28 (63.6)
2	72 (18.6)	1 (1.4)	8 (11.1)	35 (48.6)	28 (38.9)
3	71 (18.6)	3 (4.2)	4 (5.6)	26 (36.1)	39 (54.2)
4	59 (15.2)	0 (0.0)	12 (20.3)	37 (62.7)	10 (16.9)
6	69 (17.8)	1 (1.4)	8 (11.6)	42 (60.9)	18 (26.1)
7	28 (7.2)	0 (0.0)	2 (7.1)	13 (46.4)	13 (46.4)
8	41 (10.6)	0 (0.0)	3 (7.3)	13 (31.7)	25 (61.0)

The inter-rater reliability testing confirmed that there is notable variation in the overall risk scores assigned by different workers, even when given the same fictional scenario. The mean overall score was 103.7 with a standard deviation of 34.0 (Table 4). The overall ICC was low at 0.428. This scoring discrepancy, and the associated ICC values, varied across sub-categories. The Environmental

Health subcategory scores had the highest ICC (0.946), followed closely by Mental Health (0.889) and Physical Health (0.859). These values indicate a high level of agreement among different scorers, resulting in minimal measurement error for these subcategories. The Support System scores had the lowest ICC, at 0.514.

Table 4. Intraclass Correlation Coefficients (ICC), overall and for each APSART sub-score.

APSART Category	Mean Score (SD)	Intraclass Correlation Coefficient	95% CI
Overall	103.7 (34.0)	0.428	(-0.029, 0.899)
Physical Health	17.4 (7.8)	0.859	(0.567, 0.983)
Mental Health	10.6 (8.4)	0.889	(0.641, 0.987)
Cognitive Functioning	16.3 (11.1)	0.711	(0.173, 0.964)
Environmental Risks	12.6 (12.0)	0.946	(0.820, 0.994)
Support System	14.6 (10.7)	0.514	(-0.104, 0.931)
Abuse/Neglect/Exploitation Factors	18.0 (8.7)	0.686	(0.097, 0.961)
Economic Resources	9.1 (9.5)	0.612	(-0.047, 0.950)
Alleged Perpetrator	5.1 (9.8)	0.659	(0.101, 0.956)

Qualitative discussion following the IRR testing revealed specific areas of confusion that can be modified. For instance, the current Alleged Perpetrator category systematically excludes self-neglect cases. While many workers chose to skip this category for self-neglect cases, some selected the "other" field for what they judged to be the appropriate risk level. Additional discussion focused on defining certain terms (i.e. environment) and what to do if there is not enough information to assess one or more categories.[2]

Conclusions

Overall, our study revealed that the APSART score is a valid predictor of risk of a repeat APS report. However, our results are likely an underestimate of the true validity of the APSART, given the low overall inter-rater reliability. It will be important to adapt the tool and/or provide additional training to improve reliability in those areas with low ICC values. It was clear from the IRR testing that some workers provided consistently higher or lower scores; these data have been provided to the unit supervisor to address on an individual basis.

That scores are skewed so strongly toward the "high" risk level also raises the question of whether the scoring system should be recalibrated. It is possible to have a score as high as 240, but the cut-point for being categorized as "high risk" is a score of 80. The moderate risk category has a score range of 40 points, whereas the low and no risk categories each only capture a 10-point range of scores.

---

[2] As of the writing of this report, all 5 scenarios had not been fully discussed, due to time constraints. Plans are in place to complete this exercise and make appropriate changes in the near future.

It is worth considering broadening the low and moderate risk categories in order to achieve a more normal score distribution. This would mean a more sensitive tool, which would better meet the APS Unit's needs for assessing change in client risk over time.

If re-evaluating the scoring system, we should consider possible interaction between the risk categories, in order to weight different combinations of risk factors appropriately. In other words, a combination of risk factors in two or more categories may result in higher risk than would be experienced based on each of those risk categories alone, and this may vary depending on the risk factors being combined. As it stands, if an individual is considered low risk in every sub-category, his or her overall score would be 80, which is categorized as high risk. Clearly, this is an attempt to take into account the interaction issues discussed here, but it may be over-simplifying the relationships between risk factors and/or overstating the combined risk of "low" scores in all the sub-categories.

Yet despite the stated sensitivity and reliability concerns, our testing indicates that the APSART score does appear to be significantly associated with the likelihood of a repeat APS report. In other words, we can be confident that individuals with a high score are indeed at a higher risk of having a repeat report of maltreatment than individuals with a low score. However, if changes are made based on the weaknesses identified here, it will be important to re-test the validity and reliability of the tool.