Preliminary Development of a Rapid Assessment of Supervision Scale for Young Children

WHAT’S KNOWN ON THIS SUBJECT: Assessing for adequacy of supervision in the clinical setting is challenging and may result in significant variability in care. Clinicians must quickly decide if a child and family necessitate direct counseling, further intervention, or require reporting to state agencies.

WHAT THIS STUDY ADDS: This study identified the most important characteristics for the evaluation of the adequacy of supervision of a young child. A standardized scale using these characteristics may result in an efficient means to reduce variability in care.

abstract

OBJECTIVES: Standardized evaluation tools have been shown to reduce variability in care. The objective of this study was to develop a clinically oriented evaluation tool for the rapid assessment of the adequacy of supervision of a young child.

METHODS: The Rapid Assessment of Supervision Scale (RASS) was developed via a 3-step process: (1) a modified Delphi survey of child abuse experts identified the most important characteristics for use in the assessment of adequacy of supervision; (2) the RASS was designed by using standardized definitions and the results of the Delphi survey; and (3) a total of 4 medical professionals evaluated 139 real case scenarios by using the RASS. Reliability and feasibility were assessed.

RESULTS: Sixty-seven child abuse experts participated in round 2 of the Delphi process and 50 participated in round 3. The RASS included 9 supervision characteristics identified from the Delphi process, standardized definitions, and a scoring system. The interclass correlation coefficients for interrater reliability of the mean RASS scores and overall supervision classification were 0.63 (95% confidence interval: 0.56–0.70; \( P = .000 \)) and 0.59 (95% confidence interval: 0.51–0.67; \( P = .000 \)), respectively, indicating moderate to strong agreement. For intrarater reliability, correlation coefficients for mean RASS scores indicated moderate to high correlation (0.50–0.83). Correlation for overall classification of supervision ranged from low to high (0.27–0.80).

CONCLUSIONS: The RASS scale has been shown to be efficient and, in a small sample, to have moderate to substantial interrater agreement. Further development could result in a tool that aids clinicians and researchers in the evaluation of supervision. Pediatrics 2012;129: e1517–e1524
In the year 2009, >75% of child victims of maltreatment suffered neglect, and more than one-third of child deaths due to maltreatment were attributed solely to neglect, making it the single leading cause of death due to maltreatment. Supervisory neglect occurs “when a parent or caretaker fails to provide the child with adequate protection from harmful people or situations.” Younger children are generally at higher risk for injuries and other negative outcomes due to inadequate supervision. The evaluation of supervision necessitates the consideration of both direct and indirect supervision characteristics. Direct supervision may be evaluated on the basis of physical proximity to the child and an ability to take protective action based on that proximity. Indirect supervision considers situations when the caregiver knows where the child is but is unable to take immediate protective action.

Previous work has identified a consensus among parents, Children’s Protective Services (CPS) workers, and medical providers that preschool-aged children require constant supervision. However, characteristics of supervision may vary, including specifics such as caregiver distance from the child and time the child is left unsupervised. These factors may be altered based on characteristics such as environmental risks and developmental status of the child. Determining whether supervision is appropriate in any given situation is often difficult, and medical providers have identified a need for clinical tools to assist in this determination. As a result, medical providers and social workers may often make a “best guess” regarding the appropriateness of supervision of a child based on their subjective impressions of the child, family, and their environment, potentially resulting in large variability in subsequent actions (e.g., reporting to CPS, discussing directly with the family, referral to family support/education programs, doing nothing). Structured diagnostic tools and explicit criteria have been shown to reduce variability in care in mental health patients.

A diagnostic tool aimed at reducing variability in the assessment of adequacy of supervision may facilitate greater objectivity and consistency in determining which families need further intervention. Expert consensus is 1 approach to identifying important characteristics for use in the evaluation of adequacy of supervision. The Delphi method can be used to develop consensus among a group of experts and is particularly helpful in situations when subjective opinions are likely to form the primary basis for decision-making. Once components of an evaluation tool are identified, initial reliability testing of the tool is necessary by piloting among a sample of potential end users.

The purpose of this project was to develop a diagnostic tool for the rapid assessment of the adequacy of supervision of a young child during specific episodes or time periods. This tool was developed via a 3-step process: (1) a modified Delphi survey of child abuse experts was used to identify the most important characteristics for use in the assessment of adequacy of supervision; (2) the Rapid Assessment of Supervision Scale (RASS) was designed by using standardized definitions and the results of the Delphi survey; and (3) reliability was evaluated by applying the RASS to existing injury case data.

**METHODS**

**Content of RASS**

A 3-round Delphi process was used to identify and rank characteristics to be considered when assessing the adequacy of supervision in children <5 years of age. The basic principle of the Delphi process is to provide numerous experts with multiple rounds of questionnaires, along with controlled feedback from the previous responses, eventually resulting in a consensus.

Participants in the Delphi process were provided an adaptation of Dubowitz’s definition of adequate supervision as a guide: “Adequate supervision is that considering the child’s developmental level, typical behaviors, and the environment, the child is supervised in a manner that minimizes the risks of moderate or serious harm.”

**Round 1**

Round 1 of the Delphi process consisted of the identification of possible supervision characteristics that might be important for assessing the supervision of young children. A panel of medical professionals (the study authors) with expertise in child maltreatment, child injury, and supervision reviewed the published literature for previously identified characteristics. In addition, members of the panel added pertinent characteristics that have not been detailed in the literature.

Rounds 2 and 3 consisted of an electronic survey transmitted to members of the Ray Helfer Society via the society’s listserv. The Ray Helfer Society is an honorary association of physicians with expertise in child maltreatment. The e-mail sent to members invited participation in the survey briefly outlined the purpose of the project, explained that the survey was voluntary and anonymous, and included a link to the survey. The survey was conducted by using SurveyMonkey, an Internet-based tool for collecting survey data.

**Round 2**

Round 2 invited participants to review the list of characteristics (Table 1) related to adequacy of supervision identified in round 1 and numerically rank the characteristics from most important (#1) to the least important (#16). In
addition, participants were instructed to suggest any additional characteristics not already included that they thought should be considered. Participant demographic characteristics were also collected.

The ranking results of the characteristics from round 2 were evaluated by importance and consensus as described by Hecht. The importance rating is the mean of the rankings for each characteristic. The consensus rating for each characteristic is the sum of the absolute value of each respondent’s rank minus the importance rating for each characteristic. The characteristics that ranked both in the top half of consensus and bottom half of importance (high consensus of low importance) were then eliminated from consideration. Characteristics remaining after round 2 were placed on the final, round 3 survey. In addition, any participant-suggested characteristic that was mentioned on at least 10% of surveys in round 2 was also included in the round 3 survey.

### Round 3

Round 3 survey distribution, participant demographic and survey data collection, and data ranking occurred in the same manner as in round 2, with the exception that no participant-suggested comments were solicited.

### Design of RASS

The results of round 3 were considered to represent the most important characteristics that may be used to assess adequacy of supervision of children <5 years of age. These characteristics were used to create the RASS (Fig 1). Because each characteristic may represent a different level of risk in various situations, an assessment of risk for each characteristic was included. Five categories of risk of injury ranging from no risk to high risk, along with definitions, were developed based on research by Ewigman et al. Specific instructions, spaces for total and mean scores, and categories for overall classification were added to the RASS.

### Assessing the Reliability of the RASS

Scenarios describing child, parent, environment, and supervision characteristics of young children were used to assess the reliability of the RASS. These scenarios were obtained as part of the Helping Understand Risk to Toddlers (HURT) study, an ongoing case-crossover study of adult supervision of children and its relationship to injuries in children <5 years of age. Data from the HURT study were made available by the HURT principal investigator (Dr Schnitzer). The HURT data provided for this reliability assessment included de-identified child and family demographic information, household composition, neighborhood hazards, and child health/development issues. In addition, data on specific supervision characteristics were provided for 3 time periods for each child: 1 injurious episode and 2 “control” time periods when no injury occurred. However, identification of whether the scenario described resulted in injury was not included in the HURT data provided for
this analysis. These supervision data included activity of the child and parent, presence of other adults/children, supervisor distractions, time of day, proximity of supervisor, continuity of supervision, and level of supervisory attention. HURT study data used in this project were collected at the University of Missouri–Columbia from September 1, 2008, through December 31, 2009.

The 4 study authors used the RASS to evaluate 139 randomly chosen HURT supervision episodes. The authors were blinded as to whether the supervision episode resulted in injury to the child. Twelve randomly selected scenarios were repeated to evaluate test-retest reliability. Individual case scenario data were provided to the 4 raters simultaneously, in both written and oral formats, by the study coordinator. Each rater represented a different area of training/expertise: a pediatric emergency medicine physician, a child abuse pediatrician, a primary care pediatrician, and a child injury epidemiologist. Raters assessed each scenario individually and were blinded to each others’ ratings. No communication among raters regarding the rating process was allowed. Individual raters scored each individual risk characteristic and provided an overall classification, as indicated on the RASS scoring tool (Fig 1). Because HURT data were collected from real situations, not all characteristics included in the scoring tool were available in each scenario.

### Statistical Evaluation

To assess interobserver reliability, mean scores and overall classifications on the RASS for the individual cases among the 4 raters were compared by using the intraclass correlation coefficient (ICC). For categorical variables such as overall...
classification, the ICC is equivalent to a weighted $k$.\textsuperscript{15}

To assess intraobserver reliability, the test-retest method was used for 12 scenarios. The 2 mean scores and overall classification ranks for each of the 12 scenarios for each rater were evaluated by using the Pearson product-moment correlation coefficient ($r$).

The 2-way mixed effects model with single measures was used for the ICC in all instances. Two-tailed $P$ values <.05 were considered statistically significant.

To evaluate alternatives to the mean score as a functional output of the RASS, the following analysis was conducted. The number of characteristics rated $>2$ in each case was tallied and labeled “extreme ratings,” with a possible range of 0 to 9, based on the 9 characteristics potentially rated in each case (Fig 1). To evaluate if extreme ratings, rather than the mean score, were a better predictor of overall classification, the Pearson correlation coefficients ($r$) for (1) mean score and overall classification; and (2) extreme ratings and overall score were calculated. Finally, correlation coefficients were calculated to search for correlations within the individual RASS characteristics and among the individual RASS characteristics and overall rating. This study was approved by the institutional review board of the primary author’s institution.

RESULTS

Content and Design of RASS

Round 1 of the Delphi process resulted in the identification of 16 possible characteristics for consideration (Table 1). There were 67 participants in round 2 and 50 participants in round 3. The majority of participants in rounds 2 and 3 were female (77% and 68%, respectively), had parented at least 2 children (66% and 70%), and had $>10$ years of experience in child abuse medicine (60% and 58%). Table 1 includes results from the ranking of these characteristics in the Delphi process. After round 2, there were 4 characteristics eliminated based on high consensus of low importance (characteristics 4, 13, 15, and 16). In round 2, there were a total of 19 participant-suggested characteristics for possible use in evaluating supervision; however, none met the 10% threshold for inclusion in round 3. Round 3 resulted in the elimination of characteristics 1, 6, and 7, leaving 9 characteristics to be included in the RASS (Fig 1).

Assessing Reliability of the RASS

The 139 supervision episodes reviewed involved children with a mean $\pm$ SD age of 2.4 $\pm$ 1.2 years (range: 0.6–4.9 years). Figure 2 contains a case example of HURT data and ratings from the 4 raters.

Case Information Provided to Raters:
Age $=$ 2 years, 1 month
Sex $=$ Male
Household composition $=$ 1 adult, 5 children
Parent assessment of child’s health $=$ Good
Highway/main traffic artery accessible $=$ Yes, 2 lanes
Made inaccessible to kids by barrier $=$ No
Abandoned cars, buildings, or appliances in neighborhood? No
Drowning hazards? Swimming pool with barrier fence
Fall hazards $=$ 2
Animal hazards $=$ None

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
<th>Rater 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child’s age</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Child’s development/ Maturity/ Skill level</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Physical, mental, and emotional capabilities of the caregiver</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
</tr>
<tr>
<td>Responsibilities given to the child were age appropriate</td>
<td>NA</td>
<td>2</td>
<td>1</td>
<td>NA</td>
</tr>
<tr>
<td>Level of supervisory attention paid to the child (visual, auditory, neither)</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Proximity of supervising caregiver (touching, within reach, beyond reach)</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Continuity of supervision provided by the caregiver (constant or intermittent)</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Length of time child unsupervised</td>
<td>NA</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Inherent dangers in the child’s environment (swimming pools, animals, stains)</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total score</td>
<td>13</td>
<td>14</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Mean score</td>
<td>2.6</td>
<td>2.3</td>
<td>1.57</td>
<td>2.2</td>
</tr>
<tr>
<td>Overall classification</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

FIGURE 2
Case example using the RASS. NA, not available.
in the statistical analysis involving overall classification.

The interrater reliability of the mean scores for the 4 raters, as assessed by using the ICC, was 0.63 (95% confidence interval: 0.56–0.70; \( P = .000 \)), indicating moderate to strong agreement.\(^{16} \) For overall classification, the ICC was 0.59 (95% confidence interval: 0.51–0.67; \( P = .000 \)), indicative of moderate to substantial agreement.\(^{17} \) Results of the intrarater reliability assessment are displayed in Table 3. The correlation coefficients (\( r \)) for mean scores indicate moderate to high correlation.\(^{18} \) Correlation for overall classification ranged from low to high.

### Assessing Alternatives to Mean Score as Output

Correlations between mean score and overall rating (0.65) and extreme ratings and overall rating (0.61) were similar. Of the 36 possible correlations among individual RASS characteristics, 3 had \( r \) values >0.7 with statistical significance; “level of supervisory attention paid to the child,” “proximity of supervising caregiver,” and “continuity of supervision provided by the caregiver” were each independently correlated with one another.

When individual characteristics were evaluated for correlations with overall rating, the 6 highest correlations with statistical significance had \( r \) values between 0.41 and 0.75, with “responsibilities given to the child were age appropriate” having the highest correlation.

### DISCUSSION

This study identified the most important characteristics to be used for the assessment of supervision of a child <5 years of age. These characteristics have been crafted into a clinical tool for the assessment of adequacy of supervision in young children. Preliminary reliability characteristics indicate that moderate to substantial agreement among users may be attained in the assessment of supervision when using the RASS. In addition, assessing collected information with the RASS is efficient, as each assessment was completed in <2 minutes.

A wide array of clinicians, including medical professionals and social workers, evaluate children when concerns of appropriate supervision exist. Clinical decision-making in these cases is complex and may involve the relationship with the family, personal bias, and clinician characteristics, such as age, gender, upbringing, professional training, and geographic or other cultural norms. Clinicians must rapidly gather information regarding the adequacy of supervision of a child, in the context of the child’s development and environment, and make decisions regarding addressing the issue themselves or further involvement of social work and/or CPS. It is likely that, similar to child physical abuse, there is significant variability and bias in the evaluation of supervision.\(^{19–21} \) A structured scale such as the RASS may serve to reduce variability by systematizing the clinical approach.\(^{8,9} \) CPS workers may also find the RASS useful in assessing allegations of supervisory neglect, specifically those involving allegations.

### TABLE 3 Intraobserver Reliability Assessment of the RASS Using the Pearson Product-Moment Correlation Coefficient (\( r \))

<table>
<thead>
<tr>
<th>Factor</th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
<th>Rater 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( r )</td>
<td>( p )</td>
<td>( r )</td>
<td>( p )</td>
</tr>
<tr>
<td>Mean score</td>
<td>0.71</td>
<td>.009*</td>
<td>0.83</td>
<td>.001*</td>
</tr>
<tr>
<td>Overall classification</td>
<td>0.27</td>
<td>.40</td>
<td>0.80</td>
<td>.009*</td>
</tr>
</tbody>
</table>

* \( P < .05 \) refute the null hypothesis that the true correlation coefficient is 0.
of poor supervision resulting in risk of injury to the child.

Previous publications have provided guidance to clinicians regarding the assessment of supervision; however, no standards exist regarding this assessment or the specific necessary characteristics of “constant” supervision. The assessment of supervision is multifactorial, including child, caregiver, and environmental characteristics and is subjective. Due to the nature of the evaluation, the Delphi method presented the most rational process to identify the key characteristics for use in the assessment of the adequacy of supervision. In this study, the Delphi process was used to reach consensus on which characteristics are not useful or important in assessment of supervision. The remaining characteristics were then included in the RASS.

Given the variability in specific circumstances, some characteristics may be important in some instances, while relatively unimportant in others. The design of the RASS allows for clinicians to include the important characteristics in each given situation. In addition to its potential clinical uses, the RASS may be used for research purposes to characterize the variability of assessments of groups of providers, such as comparing social workers with physicians, when assessing adequacy of supervision. Further development and testing of the RASS are necessary before clinical and/or research use.

Use of the RASS to assess adequacy of supervision resulted in moderate to substantial agreement among a small group of users. Differences in evaluation may be related to raters’ training, professional or parental experience, and/or other personal characteristics. A case-based study, comparing current clinical methods with the RASS, that assesses the variability in assessment of supervision and subsequent interventions, including reporting to CPS, is necessary to determine if use of the RASS reduces variability. In most instances, the intraobserver reliability was high for both the mean score and the overall rating (Table 3). However, this was not universal in our small sample of raters, particularly with overall classification. Intraobserver reliability may be improved by linking mean score to overall rating on the RASS, such that the specific mean scores more clearly suggest particular overall ratings. Finally, based on these preliminary results, mean score seems to be a reasonable choice for functional output of the RASS; however, future studies need to examine the correlations among individual RASS characteristics and between individual characteristics and overall rating by using a larger sample size.

This study has several limitations. The results of the Delphi process may have been different if a different sample was used. Results of the study may not be generalized, given that only 4 raters from a single site used the RASS. However, this is a preliminary study and generalizability was not the intent, and future broader studies may address this issue. The RASS is not designed to assess chronic questionable supervision or predict the quality of future supervision; however, multiple individual episodes of questionable supervision may be assessed by using the RASS. Future modifications of the RASS may also need to include an explicit rating of potential severity of harm to increase validity and reliability. The impact on efficiency and scoring of the RASS when specifically collecting information in a “real-life setting” is not currently known.

CONCLUSIONS

The RASS scale allows for the rapid assessment of supervision of young children. The scale has been demonstrated to be efficient and, in a small sample, to have moderate to substantial intrarater agreement. Further development may result in a tool that aids clinicians and researchers in the evaluation of supervision and reduces variability in care.

REFERENCES

4. Anderst JD, Dowd MD. Comparative needs in child abuse education and resources: perceptions from three medical specialties. Medical Educ Online. 2010 Jul 20;15. doi: 10.3402/meo.v15i0.193


11. Hecht A. A modified Delphi technique for obtaining consensus on institutional research priorities. Paper presented at the Annual Meeting of the North Central Region AERA Special Interest Group on Community College Research; New York, NY; July 14, 1977


Preliminary Development of a Rapid Assessment of Supervision Scale for Young Children
Jim Anderst, M. Denise Dowd, Patricia Schnitzer and Tom Tryon
Pediatrics 2012;129:e1517; originally published online May 7, 2012;
DOI: 10.1542/peds.2011-2880

Updated Information & Services
including high resolution figures, can be found at:
/content/129/6/e1517.full.html

References
This article cites 10 articles, 5 of which can be accessed free at:
/content/129/6/e1517.full.html#ref-list-1

Subspecialty Collections
This article, along with others on similar topics, appears in the following collection(s):
Developmental/Behavioral Pediatrics
/cgi/collection/development.behavioral_issues_sub

Permissions & Licensing
Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
/site/misc/Permissions.xhtml

Reprints
Information about ordering reprints can be found online:
/site/misc/reprints.xhtml
Preliminary Development of a Rapid Assessment of Supervision Scale for Young Children
Jim Anderst, M. Denise Dowd, Patricia Schnitzer and Tom Tryon
Pediatrics 2012;129;e1517; originally published online May 7, 2012;
DOI: 10.1542/peds.2011-2880

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/129/6/e1517.full.html